Fortified Before Birth: Long-Run Human Capital Effects of Folic Acid Fortification

Wenjie Zhan*

Abstract

How does prenatal micronutrient exposure affect children's long-run outcomes? This paper provides new evidence from the federal mandate to fortify enriched grain products with folic acid. Exploiting cross-state variation in baseline birth defect rates linked to folate deficiency and the timing of the first trimester relative to fortification, I find that in utero exposure shifts young adults' time toward schooling: graduate school enrollment rises by 0.8 percentage points, accompanied by a 0.7-percentage-point decline in labor force participation and a $12 decline in average annual income. Effects are larger among women, White individuals, and non-Hispanic individuals. The findings are most consistent with improved math-related cognitive ability as the mechanism. Back-of-the-envelope calculations suggest that folic acid fortification generates adult earnings gains comparable in magnitude to those from other fortification programs and Food Stamps. The extremely low implementation cost of folic acid fortification implies an effectively infinite marginal value of public funds, making the policy fiscally appealing. (JEL I18, J22, J24, Q18)

Keywords: Micronutrient, Folic acid fortification, Human capital, Long-run effects

*Corresponding author at: Wenjie Zhan, University of Guelph, ON, Canada. Email: zhanw@uoguelph.ca. Address: 50 Stone Road East, Guelph, ON N1G 2W1, Canada. Wenjie Zhan is an assistant professor in the Department of Food, Agricultural and Resource Economics, University of Guelph. Supplemental analyses are available in an Online Appendix. The restricted-use Vital Statistics Natality Data can be obtained through the National Center for Health Statistics (NCHS); all other data are publicly available from ipums.org, and the websites of NCHS, Census Bureau, Burea of Economic Analysis, and the National Assessment of National Progress. The author has no conflicts of interest or funding to disclose. This study is exempt from IRB review. The author thanks Marianne Bitler, Stephen Vosti, Timothy Beatty, Richard Sexton, Rachel Soloveichik, Jenna Stearns, Ashish Shenoy, Arman Rezaee, Debora Mazetto, and Hanqi Luo for helpful comments, as well as seminar participants at UC Davis, AEA, ASHEcon, AAEA, and CAES annual meetings.

1 Introduction

The neural tube closes within the first four weeks of pregnancy—before most women know they are pregnant. During that window, maternal folate deficiency can lead to incomplete neural tube closure, causing permanent neurological damage to the developing fetus. Yet voluntary supplementation often fell short because many women were unaware that folate was needed so early in pregnancy (Petrini, Damus and Johnston, 1999). Passive, population-wide fortification therefore offered a more reliable safeguard. In March 1996, the U.S. Food and Drug Administration (FDA) mandated folic acid fortification of enriched grain products, and the immediate effect was substantial: birth defects associated with folate deficiency fell by 25–30% (Mersereau et al., 2004; Williams et al., 2015). This raises a natural question: if folate is so important for early neurodevelopment, does correcting this deficiency also translate into lasting gains in human capital?

Answering this question carries broader policy implications. First, it speaks to the design of efficient nutrition policy. If deficiency in a specific nutrient is the binding health constraint, targeted nutrient interventions may be more cost-effective than broad increases in food subsidies or medical spending. This consideration is especially pressing as government budgets face growing fiscal pressure (Bergh, 2025; International Monetary Fund, 2026). Second, mandatory folic acid fortification remains far from universal (Global Fortification Data Exchange, 2025), and the U.S. was among its earliest adopters. Evidence from the U.S. experience can therefore inform policy decisions in countries still weighing whether to mandate fortification, particularly in low-income settings where maternal folate deficiency tends to be more severe.

To estimate long-run effects of reducing folate deficiency in utero, this paper exploits two sources of variation: pre-fortification birth defect rates linked to folate deficiency across states, and the timing of the first trimester relative to the mandate. I construct these birth defect rates from birth certificate data and link them to young adults' outcomes in the American Community Survey (ACS) Public Use Microdata Sample (PUMS) by state and year of birth. Sections 4.1 and 5.1 validate baseline birth defect rates as a credible pre-period proxy for folate deficiency. Exposed cohorts are those whose first trimester oc-

curred after fortification, since this is the critical window during which folic acid reduces the risk of folate-sensitive birth defects. I use the mandate authorization date (March 5, 1996) rather than the full implementation date (January 1, 1998) as the event date, because prior work shows fortification was largely completed before the implementation deadline (Jacques et al., 1999) and the first-stage evidence in Section 5.1 is consistent with this timing. I then estimate the mandate's effects by comparing exposed and unexposed cohorts across states with differing baseline prevalence of these defects.

Results show that in utero exposure to fortification raises graduate school enrollment among young adults aged 23–30 by 0.8 percentage points (14.4% at the sample mean), accompanied by a 0.7-percentage-point decline in labor force participation (0.8% at the mean) and a $12 decline in average annual income (3.4% at the mean). Effects are largest among women, White, and non-Hispanic individuals; women also adjust on the intensive margin, with exposed women less likely to work full time. Effects for Hispanic individuals are smaller and less precisely estimated, consistent with the 1996 mandate's exclusion of some staples commonly consumed in Hispanic households.

I discuss four potential mechanisms behind this shift toward schooling. First, fertility selection is unlikely to explain the results because birth outcomes and maternal characteristics do not respond to in utero fortification exposure, aside from a small increase in the share of births to younger mothers after fortification. Second, there is no evidence that effects of in utero fortification exposure are driven by a lower likelihood of functional difficulty in young adulthood. Third, changes in local labor market conditions do not appear to drive the results. The effects are not diminished by place-of-residence-by-year fixed effects, and in utero fortification exposure does not affect unemployment or occupational quality among those who are employed. Finally, consistent with improved cognitive ability as a plausible mechanism, I find suggestive evidence that fortification exposure is associated with higher math scores among 4th and 8th graders.

Finally, back-of-the-envelope cost-benefit calculations suggest that folic acid fortification generates about $2,911 in additional annual adult income. This gain is smaller than, but still comparable in magnitude to, estimated gains from other fortification programs

and Food Stamps. Because the cost of fortification is extremely low relative to the additional tax revenue generated by these income gains, the implied marginal value of public funds (MVPF) is effectively infinite, suggesting that the policy perfectly pays for itself.

This paper contributes to food fortification literature by studying a nutrient whose relevance is concentrated at the very beginning of fetal development. Most existing economic work examines salt iodization, documenting long-run benefits for cognitive development, health, and socioeconomic outcomes (Feyrer, Politi and Weil, 2017; Serena, 2019; Adhvaryu et al., 2020; Huang, Liu and Zhou, 2020; Deng and Lindeboom, 2022*a*; Tafesse, 2022). A notable exception is Niemesh (2015), which shows that iron fortification of bread raised adult incomes and improved children's school enrollment and long-run wages. These studies provide important evidence that correcting micronutrient deficiencies can generate lasting economic gains. Folic acid fortification extends this literature to a different biological margin: folate is especially important in the earliest weeks of gestation, when the neural tube closes—often before a woman knows she is pregnant—and deficiency during this short window can cause permanent neurological damage to the developing fetus (Irvine et al., 2022). Outside economics, the epidemiological literature has focused mainly on the short-run health effects of folic acid supplementation (e.g., Wald et al., 2001; Quinlivan et al., 2002; Kancherla et al., 2022) and on cost-benefit analyses of fortification programs (e.g., Grosse et al., 2005; Bentley et al., 2009). What remains largely unknown is whether reducing folate deficiency through population-wide fortification generates lasting human capital gains. I address this gap by studying how correcting a micronutrient deficiency during the earliest stage of fetal development affects education and labor market outcomes in young adulthood.

More broadly, this paper relates to the fetal origins literature, which documents long-run consequences of prenatal and early childhood nutritional conditions. Adverse shocks such as famine (Almond et al., 2007; Chen and Zhou, 2007; Lindeboom, Portrait and Van den Berg, 2010; Scholte, Van Den Berg and Lindeboom, 2015; Deng and Lindeboom, 2022*b*) and Ramadan fasting (Almond and Mazumder, 2011; Almond, Mazumder and Van Ewijk, 2015; Majid, 2015) have been linked to poorer health and labor outcomes

in adulthood, while positive interventions such as breastfeeding (Fitzsimons and Vera-Hernández, 2022), iodine supplementation (Field, Robles and Torero, 2009; Araújo, Carrillo and Sampaio, 2021), and food assistance programs (Hoynes, Page and Stevens, 2011; Rossin-Slater, 2013; Hoynes, Schanzenbach and Almond, 2016; Bailey et al., 2024) have been shown to improve long-term socioeconomic outcomes. This paper extends that literature by studying a previously underexamined intervention targeting a nutritional deficiency that can disrupt fetal development at the earliest stage of gestation.

Finally, the paper connects the fetal origins literature to work on young adults' choices between further schooling and labor market entry. Existing studies show that post-baccalaureate schooling decisions respond to financing constraints, labor market conditions, and expected returns (Weiler, 1994; Bedard and Herman, 2008; Johnson, 2013; Altonji, Arcidiacono and Maurel, 2016; Black, Turner and Denning, 2023), and that strong labor market opportunities can pull young people away from college and into employment (Black, McKinnish and Sanders, 2005; Charles, Hurst and Notowidigdo, 2018). I add a new dimension to this literature by showing that these early adult choices may also be shaped by conditions much earlier in life. In particular, prenatal nutrition may affect not only completed human capital, but also the timing and direction of young adults' transition between additional schooling and work.

The paper is organized as follows. Section 2 provides policy background. Section 3 describes the data. Section 4 outlines the empirical methods. Section 5 presents the results. Section 6 calculates MVPFs. Section 7 concludes.

2 Background

2.1 Folate deficiency disorder and associated birth defects

Folate deficiency is a major cause of neural tube defects (NTDs), the most common congenital central nervous system anomalies (CNSAs) in newborns (Smithells et al., 1983). Severe NTDs, such as anencephaly, are typically fatal, with most affected infants dying

before or shortly after birth.[1] Infants with less severe NTDs, such as spina bifida, can survive into adulthood but often face a high risk of lifelong physical and cognitive disabilities (Yi et al., 2011).[2] In the early 1990s, roughly 4,000 fetuses in the U.S. (about 1 in 1,000) were affected by NTDs each year, with about one third lost to selective or spontaneous abortion (Cragan et al., 1995; Mersereau et al., 2004). Folate deficiency is also associated with other congenital CNSAs, such as hydrocephaly (Naz et al., 2016; Liu et al., 2018). These defects can arise as early as the first month of pregnancy, when the neural tube begins to form, and failure of the neural tube to close by the end of the first trimester can cause irreversible damage to the central nervous system (Obeid, Holzgreve and Pietrzik, 2013). While in utero surgery may provide some palliative benefits, it cannot reverse the underlying neurological damage (Greene and Copp, 2014). Timely medical intervention is often difficult because routine screening for neural tube defects typically relies on the second-trimester anatomic ultrasound at 18 to 22 weeks, and detection rates are much lower in the first trimester (American College of Obstetricians and Gynecologists, 2017). Many pregnant women in the U.S. also lack adequate prenatal care (Blumenfeld, Siegler and Bronshtein, 1993).

2.2 Sources of folate

Folate occurs naturally in foods such as beef liver, dark green leafy vegetables, beans, peas, nuts, and many fruits and fruit juices. However, meeting recommended intake during pregnancy through diet alone is difficult (Czeizel, 2000). Data from the National Health and Nutrition Examination Survey (NHANES) III (1988–1994) show that women aged 15–49 consumed an average of 233.68 μg of folate per day, well below the 400 μg recommended by the U.S. Public Health Service for pregnant women. One reason dietary intake often falls short is that naturally occurring food folate is unstable under typical cooking conditions, which can substantially reduce the amount ultimately absorbed, making diet alone a less reliable way to improve folate status during pregnancy (McNulty and Pentieva, 2004).

[1] Infants with anencephaly are born without parts of the skull and brain.

[2] In spina bifida, the backbone does not close properly, leaving a section of the spinal cord and spinal nerves exposed without full protection.

Folate is also available through supplements, including over-the-counter folic acid tablets and multivitamins. Folic acid is often recommended during prenatal visits (Ray, Singh and Burrows, 2004). A key challenge, however, is low awareness of and adherence to supplementation guidance (Toivonen et al., 2018). CDC recommendations emphasize starting folic acid at least one month before conception.[3] Yet about 50% of U.S. pregnancies are unintended (Finer and Zolna, 2016). From 1995 to 1998, only about 30% of U.S. women reported taking a daily vitamin containing folic acid, and fewer than 10% knew it should be taken before pregnancy (Petrini, Damus and Johnston, 1999). Access and affordability can also be barriers, particularly for low-income women (Czeizel, 2000). These constraints highlight the need for a low-cost, preferably passive approach to ensuring adequate folic acid intake among women who may become pregnant.

2.3 Folic acid fortification and other fortifications in the U.S.

The U.S. has a long history of using food fortification to improve public health. Salt iodization began in the 1920s, vitamin D fortification of milk followed in the 1930s, and flour and bread were enriched with B vitamins and iron in the 1930s and 1940s. The most recent effort, folic acid fortification of grain products, began in the 1990s. The first wave of grain fortification in the 1940s followed the identification of specific nutrient deficiency disorders among the U.S. population. In the early 1940s, the FDA issued the first standard of identity for enriched flour, requiring the addition of iron and B vitamins (niacin, thiamin, and riboflavin). By the 1950s, these standards extended to other cereal grain products, including bread, rice, macaroni, and noodles (Hutt, 1984; Committee on Use of Dietary Reference Intakes in Nutrition Labeling, 2004). Folic acid fortification was the most recent amendment to the standard of identity for enriched grain products and is widely regarded as one of the most successful public health initiatives in recent decades (Berry, Mulinare and Hamner, 2010).

As with earlier fortification campaigns, the decision to fortify foods with folic acid was driven by accumulating evidence that folic acid prevents neural tube defects (NTDs). In October 1990, as part of the Nutrition Labeling and Education Act, Congress directed the

[3] See https://www.cdc.gov/ncbddd/folicacid/recommendations.html (accessed on 05/20/2022).

FDA to evaluate the link between folic acid and NTDs and to develop a plan for adding folic acid to foods (Wright, 2003). On September 14, 1992, the U.S. Public Health Service (USPHS) recommended that all women of childbearing age consume 400 μg of folic acid daily to prevent NTDs. In response, on March 5, 1996, the FDA amended the standard of identity to require 140 μg/100 g of folic acid in enriched grain products by January 1, 1998 (Food and Drug Administration, 1996). In practice, fortification was largely completed by mid-1997 (Jacques et al., 1999). I therefore define the event time as March 1996, when the FDA authorized folic acid fortification. Because enriched grain products are ingredients in many processed foods, the added folic acid shows up in products beyond the enriched grain category; for example, some chips contain folic acid (Figure 1). Before the mandate, voluntary addition of folic acid was prohibited in standardized foods[4] and discouraged in other products, primarily to avoid overfortification and nutrient imbalances (Food and Drug Administration, 1996, 2015).

Figure 1: Chips with enriched wheat flour as an ingredient

3 Data

3.1 Birth certificate data

The Vital Statistics Natality files (birth certificate data) cover all U.S. live births and report detailed information on birth outcomes, including birth month and year, state of

[4] "Standardized foods" are products with a federal standard of identity, such as enriched grain products.

birth, birth weight, gestational age in weeks, and congenital anomalies, as well as maternal characteristics such as age, race, Hispanic origin, education, and prenatal care use (National Center for Health Statistics, 2003). I use these data for two purposes. First, I proxy baseline folate deficiency using the pre-fortification prevalence of folate-sensitive congenital anomalies and assign cohort-level in utero exposure using gestational age and the policy's authorization date (Section 4.1). Second, I test whether fortification exposure affects birth outcomes or shifts the distribution of maternal characteristics, which helps assess whether fertility selection could be driving the main results.

3.2 Young adult outcome data

Outcome data for young adults come from the American Community Survey Public-Use Microdata Sample (ACS PUMS), 2019–2024 (Ruggles et al., 2025). For educational outcomes, I study graduate school enrollment (including master's, professional, and doctoral programs) and educational attainment, including completion of high school (diploma or GED), and bachelor's degrees. For labor market outcomes, I examine labor force participation, full-time status, and income. These outcomes are standard measures of human capital and are commonly examined in related work (Niemesh, 2015; Adhvaryu et al., 2020).

I use the 2019–2024 PUMSs so that each survey year includes both exposed and unexposed cohorts. I focus on young adults because the oldest exposed cohorts are only 28 years old during this period. The main analysis sample includes individuals aged 23–30, which captures early-career outcomes for young adults beyond typical college ages. I discuss alternative sample definitions in Section 5.5.

3.3 Other data

I collect the following supplementary data to support my main analysis.

Baseline state characteristics. I examine how the exposure measure relates to baseline state characteristics in Section 4.1. I compile these characteristics from multiple sources. Demographic measures, including race, gender, age, and total population, come from the

Intercensal Population Estimates (US Census Bureau, 1990). Birth and death rates, the unemployment rate, the value of products sold per farm, and average farm size come from the County and City Data Book (1988) (US Census Bureau, 2009). Measures of transfer payments come from the Bureau of Economic Analysis's Regional Economic Information System (REIS) (Bureau of Economic Analysis, 1988).

Dietary and blood folate data. In Section 5.1, I examine trends in dietary folate intake and blood folate concentrations using publicly available data from NHANES III (1988–1994) and the continuous NHANES (1999–2006). NHANES III is a nationally representative survey of the civilian, noninstitutionalized U.S. population aged 2 months and older. Beginning in 1999, NHANES shifted to a continuous design, collecting data annually and releasing them in two-year cycles. Both NHANES III and the continuous NHANES combine an in-home interview with a standardized physical examination in mobile examination centers, along with laboratory testing of biospecimens. The surveys provide detailed information on demographics, socioeconomic characteristics, health conditions, and diet, as well as anthropometrics and biomarkers. For folate status, NHANES reports dietary folate intake based on 24-hour dietary recalls and measures serum and red blood cell (RBC) folate in the laboratory component. I restrict the continuous NHANES sample to 1999–2006 because folate laboratory methods changed after 2006, limiting comparability across periods (Yetley et al., 2011).

Employment data. In Section 5.5, as a robustness check, I control for a Bartik-style predicted unemployment measure to proxy for economic conditions at birth. Baseline sectoral employment shares come from the Bureau of Labor Statistics' Quarterly Census of Employment and Wages (QCEW) (US Bureau of Labor Statistics, 1989-2002). Following East (2020), for each state of birth I interact these baseline employment shares with annual national changes in sector-specific unemployment rates and then sum across sectors to construct a predicted state unemployment rate.

Test score data. In Section 5.6, I examine the association between fortification exposure and academic performance in elementary and middle school. Test score outcomes come from the National Assessment of Educational Progress (NAEP) public state-level

aggregate data for 4th and 8th graders. The public NAEP reports state average scores and the shares of students at or above the NAEP achievement levels (Basic and Proficient)[5] in mathematics, reading, science, and writing, typically at intervals of one or two years since 1990. State average scores are computed by averaging scaled scores from a representative sample of students in each state, with scores ranging from 0 to 500. I focus on 4th and 8th grade math and reading because public series for science, writing, and all 12th grade tests are available for far fewer years. The analysis sample includes 4th grade math in 2000, 2003, 2005, 2007, 2009, and 2011; 4th grade reading in 2002, 2003, 2005, 2007, 2009, and 2011; 8th grade math in 2000, 2003, 2005, 2007, 2009, 2011, and 2015; and 8th grade reading in 2002, 2003, 2005, 2007, 2009, 2011, and 2015.

4 Methods

The identifying variation comes from two sources: (i) cross-state differences in the pre-fortification prevalence of birth defects linked to folate deficiency, and (ii) differences across cohorts in whether the first trimester occurred before or after fortification. The logic is that, because folate deficiency during the first trimester can cause irreversible fetal damage, improvements in maternal folate status during this window may have lasting effects on human capital. The design is valid if pre-fortification defect prevalence captures cross-state differences in the potential to benefit from fortification, and if fortification meaningfully increased folate intake rather than simply changing policy on paper. Section 4.1 discusses the validity of using birth defects to measure cross-state exposure. Section 5.1 presents evidence that fortification increased folate intake and reduced birth defects. Relative to a randomized trial that assigns folic acid supplements to mothers and follows their children into adulthood, this design is more feasible at scale and avoids ethical concerns.

This design resembles an established approach in the literature that uses baseline regional disease prevalence to infer the benefits of a health intervention. For example, researchers use baseline hookworm infection rates to study hookworm eradication (Bleak-

[5]See https://nces.ed.gov/nationsreportcard/mathematics/achieve.aspx for details on the NAEP achievement levels.

ley, 2007), malaria prevalence to evaluate malaria eradication (Bleakley, 2010; Kuecken, Thuilliez and Valfort, 2021), measles incidence to assess vaccination (Atwood, 2022), pneumonia rates to examine the introduction of sulfa antibiotics (Lazuka, 2020), and goiter prevalence to analyze salt iodization (Feyrer, Politi and Weil, 2017; Adhvaryu et al., 2020).

4.1 Exposure measure

From the birth certificate data, I construct each state's pre-fortification prevalence of central nervous system anomalies (CNSAs), comprising spina bifida, anencephaly, hydrocephaly, microcephaly, and a residual "other" category. Folate deficiency is a leading cause of neural tube defects (NTDs) (Wald et al., 2001), the most common of which are spina bifida and anencephaly; other NTD subtypes are captured in the "other" CNSA category. Hydrocephaly and microcephaly are included because folate deficiency may contribute to these conditions as well, either directly or through NTDs (Abdel-Salam and Czeizel, 2000; Naz et al., 2016; Liu et al., 2018).

I define the baseline period as January 1989 through June 1993. Most states began reporting congenital anomalies in 1989; the exceptions are Louisiana (1990), Nebraska (1990), Oklahoma (1991), New York (1993), and New Mexico (not reported during the study period). Starting the baseline in 1989 therefore maximizes geographic coverage, encompassing all states and the District of Columbia except New Mexico. I end the baseline in June 1993 so that cohorts born afterward have at least four pre-periods available for the event-study analysis, while also ensuring that the exposure measure is predetermined relative to the fortification mandate for both exposed and unexposed cohorts. Section 5.5 shows that extending the baseline to the full pre-fortification period (1989–1995) does not noticeably change the results.

Figure 2a shows baseline CNSA rates across states. I define high-exposure states as those in the top quartile of baseline CNSA rates, yielding 14 states: IN, IA, KS, MD, MN, NE, NJ, NY, ND, RI, SD, TN, TX, and VT. I test robustness to using the continuous CNSA rate as the exposure measure in Section 5.3 and to alternative high-exposure thresholds in Section 5.5.

To validate the exposure measure, I link state-level pre-fortification CNSA rates to biomarkers of folate status from NHANES III. Table 1 shows a negative association: on average, individuals living in high-baseline-CNSA states have lower blood folate concentrations. In Section 5.1, I also show that high-baseline-CNSA states experienced larger post-fortification declines in CNSA rates (Figure 4d), which further supports the validity of the exposure measure.

Table 1: Correlation between baseline CNSA rate and folate biomarkers

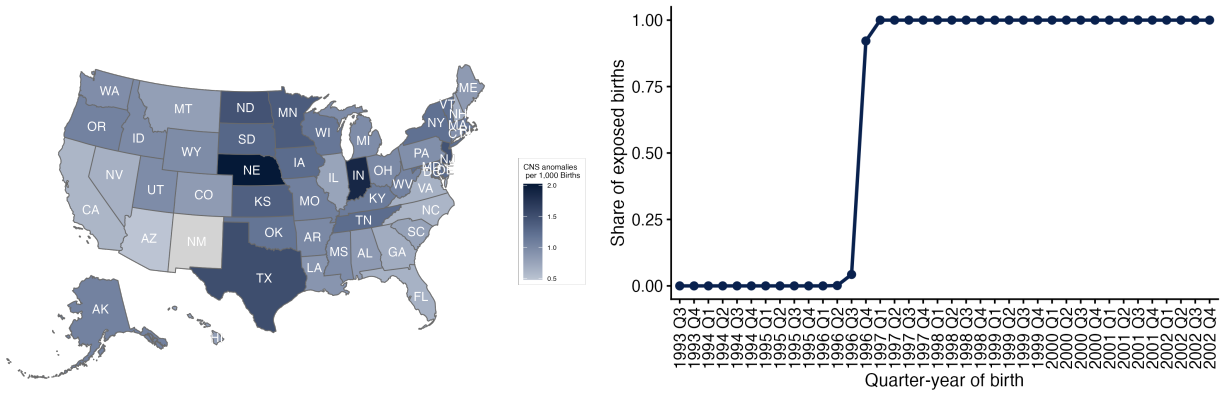
	Serum folate (1)	RBC folate (2)	Serum folate (3)	RBC folate (4)
Continuous CNSA rate	-0.538**	-11.410**		
	(0.261)	(4.572)		
High CNSA			-0.606***	-10.554***
			(0.185)	(3.171)
Observations	10,842	10,913	10,842	10,913
R^2	0.001	0.001	0.002	0.003
Dep. var. mean	7.332	198.186	7.332	198.186

Notes: Blood folate is measured in μg/mL. Robust standard errors are in parentheses. Regressions are weighted by MEC examination weights. Source: NHANES III public-use data covering 35 counties across 13 states. CNSA rate is measured at state level. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

Another concern is that baseline CNSA rates are not randomly assigned and may simply capture other state characteristics, such as differences in demographics, economic conditions, or local policy environments. For example, if states with higher baseline CNSA rates also tend to be poorer, and early-life economic conditions independently shape adult outcomes, the estimates could reflect underlying economic hardship rather than fortification. To mitigate this concern, Table A1 regresses the high-baseline-CNSA indicator on pre-1989 local characteristics, including demographic, economic, and agricultural measures at both the state and county levels. Column (1) shows that individual state characteristics have imprecise coefficients, with the full set of state-level observables explaining about 30% of cross-state variation. Column (2) shows that when all regressors are defined at the county level, these same observables explain much less—only about 5%. Adding state fixed effects raises the R^2 to roughly 30% (Column (3)). These results suggest that high-baseline-CNSA status does not simply proxy for a single observable

factor or a combination of local characteristics.

I determine exposure timing using gestational age recorded on birth certificates and map it to quarter of birth, since PUMS does not report gestational length. An infant is classified as exposed if the first trimester ends after March 1996, the month mandatory fortification took effect—neural tube closure occurs by the end of the first trimester, so folic acid can reduce CNSAs only if adequate levels are present during this window. I aggregate this birth-level indicator to the quarter-by-year level to match the PUMS data. As shown in Figure 2b, the share exposed rises sharply beginning with births in 1996 Q4, and I accordingly define individuals born in 1996 Q4 or later as the exposed cohorts. This pre–post difference, combined with cross-state variation in baseline CNSA rates, provides key identifying variations in my empirical strategy.



(a) Baseline CNSA rates by state

(b) Share of exposed births by quarter-and-year-of-birth

Figure 2: Spatial and temporal variation in fortification exposure

Notes: Figure 2a maps baseline CNSA rates (January 1989–June 1993) by state of birth. Figure 2b defines a birth as exposed if the first trimester ends after March 1996, when fortification was authorized, and then aggregates this exposure indicator to county-by-quarter-by-year averages.

4.2 Empirical method

This section discusses the empirical method, including model specifications, identifying assumptions, interpretation of the estimates, and the choice of preferred specification.

Estimation model. I estimate the following cohort difference-in-differences (DiD)

model for the baseline analysis:

$$Y_{istc} = \beta\big(\text{High CNSA}_s \times \text{Exposed cohorts}_t\big) + \gamma_{sc} + \delta_t + X'_{istc}\theta + C'_{sc}\pi + \varepsilon_{istc}, \qquad (1)$$

where Y_{istc} denotes the outcome for individual i born in state s and birth cohort t (quarter-by-year), observed in ACS survey year c. High CNSA_s is an indicator equal to 1 for states in the top quartile of the baseline CNSA rate and 0 otherwise. Exposed cohorts_t is an indicator equal to 1 for cohorts born in 1996 Q4 or later and 0 otherwise. γ_{sc} are state-of-birth-by-survey-year fixed effects, which absorb time-varying state factors common to all individuals born in the same state and observed in the same survey year. δ_t are quarter-by-year-of-birth fixed effects that control for cohort-specific shocks. X_{istc} are individual controls (gender, race indicators for White, Black, and other, and an indicator for Hispanic). C_{sc} includes controls for potentially confounding policies measured at the state-by-cohort level, including Medicaid and CHIP eligibility expansions, welfare reform, and state mental health parity laws. The error term is ε_{istc}. The coefficient β is the primary parameter of interest and captures differential changes in outcomes for exposed cohorts born in high-baseline-CNSA states relative to other cohorts and states.

Identifying assumptions. The validity of this empirical strategy relies on two assumptions: parallel trends and no anticipation. The parallel trends assumption requires that, absent fortification, average outcomes for young adults in high- and low-baseline-CNSA states would have evolved similarly across birth cohorts. To assess pre-treatment trends, I estimate dynamic effects for key outcomes using an event-study design. Specifically, I replace $\beta\big(\text{High CNSA}s \times \text{Exposed cohorts}t\big)$ with $\sum \tau\beta\tau, , \tau \neq 1995\big(\text{High CNSA}_s \times \mathbf{1}\text{Cohort} = \tau\big)$, where τ indexes exposure year and I construct exposure years by grouping cohorts into four-quarter bins after fortification. For example, exposure year 1996 includes cohorts born in 1996Q4–1997Q3. Section 5.5 further assesses the credibility of these event-study results by examining sensitivity to hypothesized violations of the parallel trends assumption following Rambachan and Roth (2023).

The no-anticipation assumption requires that (i) prospective mothers did not change behavior in anticipation of fortification, and (ii) food manufacturers did not begin for-

tifying before March 1996. In this setting, anticipatory responses are unlikely. Petrini, Damus and Johnston (1999) documents low awareness of folic acid among women of childbearing age, and because the mandate was motivated by scientific evidence and directed at food manufacturers, it likely had low salience for the general public. On the supply side, voluntary folic acid fortification was prohibited for standardized foods and discouraged elsewhere due to concerns about overfortification and nutrient imbalances (Food and Drug Administration, 1996, 2015). Consistent with these arguments, the event-study results show no evidence of anticipatory behavior.

Interpretation. Because the threshold at which fortification would have zero effect is unknown, low-baseline-CNSA states may also have benefited. As a result, the cohort-DiD design does not compare treated and untreated groups. Instead, it identifies a relative effect: the impact in top-quartile states relative to the average impact in all other states. If fortification improved outcomes outside the top quartile, this contrast understates the effect in high-baseline states and does not recover the nationwide average effect. The estimates should therefore be interpreted as a lower bound for high-baseline states.

Binary treatment versus continuous treatment. The baseline specification compares states in the top quartile of baseline CNSA exposure with all other states. Although some prior studies use dose-response specifications with continuous treatment as their main approach (Niemesh, 2015; Adhvaryu et al., 2020), I use a binary specification for baseline results for two reasons.

First, identifying the average causal response with a continuous treatment requires a stronger parallel trends assumption, which may be implausible in this setting. Let $Y_t(d)$ denote the potential outcome at time t under treatment dose d. Callaway, Goodman-Bacon and Sant'Anna (2025) (hereafter CGS25)[6] show that TWFE estimates with a continuous treatment, as in existing fortification papers (Niemesh, 2015; Adhvaryu et al., 2020), should be interpreted as the average causal response to treatment dose. With heteroge-

[6]Callaway, Goodman-Bacon and Sant'Anna (2025) was last revised on December 31, 2025. As of April 2026, it is conditionally accepted at the AER, so this version should be reasonably close to the final published version. A related paper is Callaway, Goodman-Bacon and Sant'Anna (2024*a*), which has circulated as an NBER working paper. However, the assumptions and estimation procedures differ in important ways across the two papers. I therefore focus on Callaway, Goodman-Bacon and Sant'Anna (2025).

neous effects across doses, this interpretation relies on what they call the Strong Parallel Trends (SPT) assumption: the average evolution of outcomes for the entire treated population, had all units experienced dose d, must equal the trajectory actually experienced by the group with dose d. This is stronger than the standard parallel trends (PT) assumption in the binary specification, which requires comparable trends across two potential outcome paths, $Y_t(H)$ for high-exposure states and $Y_t(L)$ for lower-exposure states. The SPT assumption instead requires comparable trends across the full distribution of potential outcome paths, one for each dose level $Y_t(d)$.

CGS25 caution against adopting this assumption uncritically, noting that it may be implausible in many applications (CGS25, p. 13). In this setting, it is more plausible that cohorts born in high-exposure states would have followed outcome trajectories parallel to those born in lower-exposure states than that cohorts across the full distribution of baseline CNSA rates would have satisfied the set of parallel trajectory restrictions required by SPT. This concern is especially relevant if treatment effects vary across dose levels. TWFE estimates with continuous treatment also have difficult-to-interpret weighting schemes. CGS25 propose a nonparametric alternative that avoids these weighting problems, but its identification still relies on the SPT assumption. In Section 5.3, I provide evidence that SPT is unlikely to hold in my setting.

Second, the binary specification is less sensitive to misspecification and measurement error in the exposure measure. If the relationship between baseline CNSA rates and true exposure is nonlinear, or if baseline CNSA rates are measured with error, states are more likely to be classified correctly into broad exposure groups than to be placed correctly throughout the continuous distribution. The quartile-based specification is therefore more robust to small errors in measured exposure levels and to misspecification in how baseline CNSA rates map into actual fortification exposure.

Despite these concerns, Section 5.3 estimates the dose-response relationship using continuous CNSA rates following CGS25 and finds results that are broadly consistent with the baseline estimates. Because of the identification concerns discussed above, however, I treat the dose-response estimates as complementary evidence rather than the primary

basis for inference.

5 Results

5.1 First stage

This section presents first-stage evidence on the effects of fortification. Descriptive evidence suggests that, after fortification, folic acid content in foods increased, dietary folate intake and blood folate concentration rose, and the CNSA rate declined.

Folate content in foods increased after folic acid fortification. First, folic acid content in foods increased sharply after fortification. Using the 1996 and 1998 Continuing Survey of Food Intakes by Individuals (CSFII), Lewis et al. (1999) compare per-serving folic acid in the same foods before and after the mandate, based on USDA recipe-based nutrient calculations. Because the CSFII records the reason for composition changes, including enrichment/fortification, reformulation, agricultural or processing modifications, and implementation of the Nutrition Labeling and Education Act, they can isolate changes attributable to fortification. As shown in Figure 3, folic acid increased across a wide range of products, from white bread to snack and cookie bars; in total, more than 350 basic food items show higher folic acid due to fortification (Anderson et al., 2001).

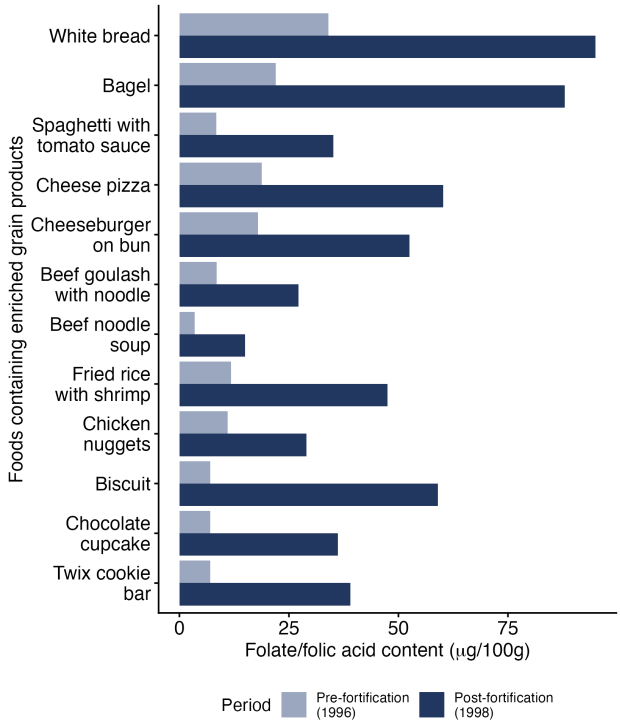


Figure 3: Changes in folate contents in selected foods attributable to fortification

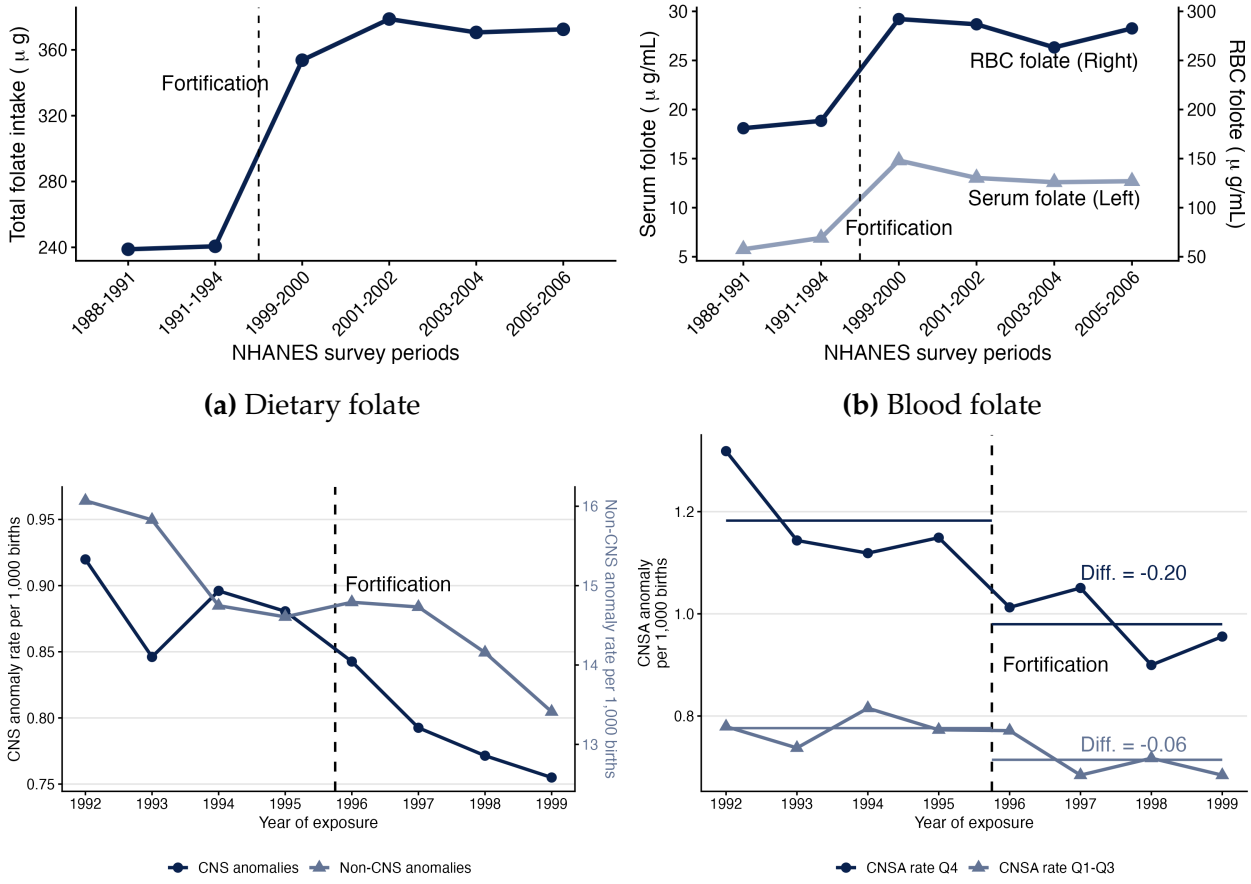
Notes: This figure uses estimates from Lewis et al. (1999). Food folate content data come from the USDA Continuing Survey of Food Intakes by Individuals (CSFII) 1996 and 1998. Folate content is estimated by the USDA using recipe-based calculations. The changes shown reflect only the portion attributable to fortification.

Dietary folate intake and blood folate increased after folic acid fortification. Second, using NHANES, I document sharp post-fortification increases in dietary folate intake and blood folate among women aged 19–45. Mean dietary intake rose by nearly 50% (Figure 4a). The share with intake below 400 μg/day fell from 98.65% to 69.87% ($\Delta = -28.78$ percentage points). These intake measures exclude folic acid from supplements and medications (Ahluwalia et al., 2016). Biomarkers show similar gains: serum folate more than doubled and red blood cell (RBC) folate rose by nearly 50% (Figure 4b), indicating sustained improvements in folate status.[7]

Congenital anomalies declined after folic acid fortification. Finally, Figure 4c shows that CNSA rates declined as folate intake and absorption increased. After a relatively flat

[7]Laboratory methods for blood folate were stable from 1999 to 2006 but changed after 2006 (Yetley et al., 2011; Pfeiffer et al., 2012). Because public-use NHANES does not report state identifiers, I cannot separately plot trends for high- and low-baseline-CNSA states.

period from 1992 to 1996, incidence fell substantially following fortification. That rates of other congenital anomalies were stable over the same period and did not begin to decline until after 1997 suggests this pattern is unlikely to reflect broad improvements in health. Figure 4d shows larger post-fortification declines in states with higher baseline CNSA rates, supporting the use of the high-baseline-CNSA indicator as the exposure measure in my research design.



(a) Dietary folate

(b) Blood folate

(c) CNSA and non-CNSA rates

(d) CNSA rates in high- and low-baseline-CNSA states

Figure 4: Trends in dietary folate, blood folate concentration, and CNSA rate

Notes: Data in Figures 4a and 4b come from the harmonized NHANES files constructed by Nguyen et al. (2023), which ensure comparability of folate measures across survey periods. All folate measures use Mobile Examination Center (MEC) final examination sample weights. Data in Figures 4c and 4d come from the Vital Statistics Natality files (birth certificate data).

5.2 Main results

Figures 5a–5f report the effects of in utero exposure to folic acid fortification on educational and labor market outcomes for young adults aged 23–30 across birth cohorts. These results tell a consistent story: fortification exposure shifts time away from work and toward schooling during early adulthood.

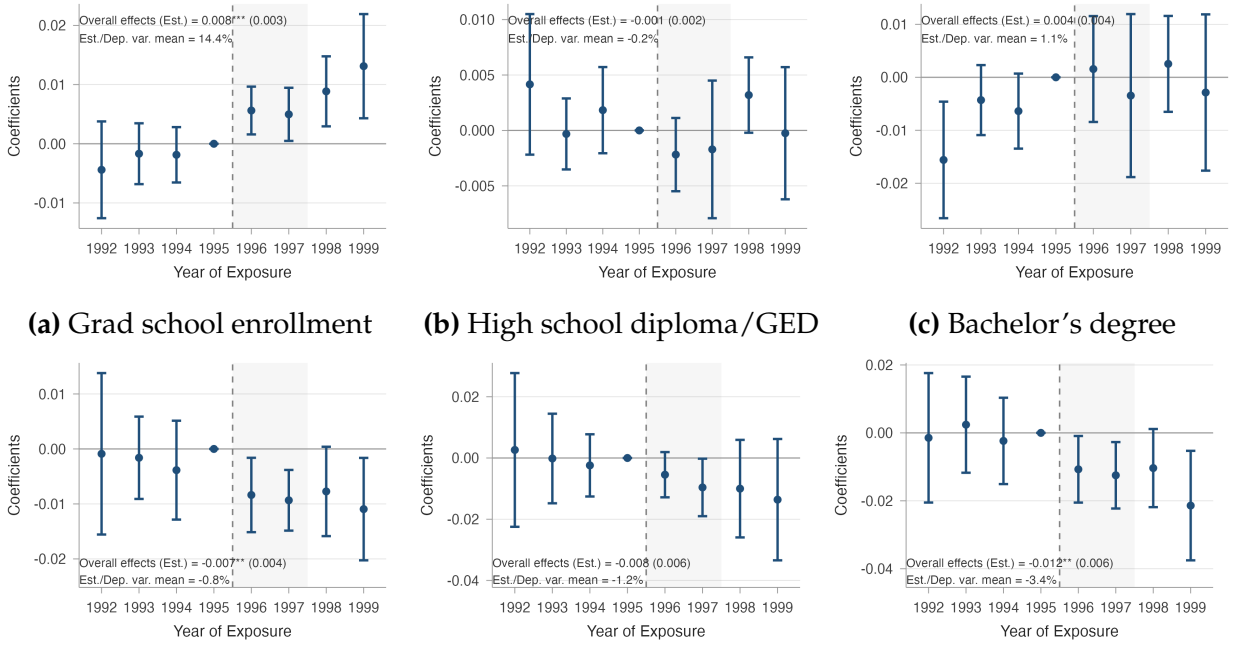
Figures 5a–5c show that graduate school enrollment rises by 0.8 percentage points (14.4% at the mean), while high school diploma or GED completion and bachelor's degree attainment change little. The absence of effects at lower education margins suggests that fortification mainly affects individuals who would have completed high school and earned a bachelor's degree regardless, but who otherwise would not have continued to graduate school. These findings are consistent with Niemesh (2015), who find no effects of iron fortification of bread on high school completion or college attendance among adults aged 23–30.

The null effects on high school completion and bachelor's degree attainment may have several explanations. First, high school completion rates are already above 90% even in exposed states, leaving limited room for improvement. Second, the remaining gaps in high school completion and bachelor's degree attainment may not be driven mainly by lower cognitive ability. For example, some young adults may choose not to attend or complete college because of financial constraints or because they pursue careers that do not typically require a college degree. Third, as discussed in Section 5.6, fortification is associated with a small increase in the share of births to younger mothers. If these marginal births tend to have worse outcomes, compositional changes could partly offset gains among individuals who would have been born regardless.

The labor market results follow directly from the increase in graduate school enrollment. Figures 5d–5f show a 0.7-percentage-point decline in labor force participation (0.8% at the mean) and a $12 decline in annual income (3.4% at the mean). The income effect is modest for several reasons: the sample includes all young adults, not only the employed; the income measure is net income, which incorporates financial losses and can be neg-

ative; and roughly 20% of employed individuals in the sample work part-time— fewer than 35 hours per week. The schooling and labor force participation effects are similar in magnitude, consistent with a reallocation of time across these two margins rather than independent adjustments to each.

These event-study estimates show no clear evidence of pretrends. Section 5.5 further examines the sensitivity of these results to potential violations of the parallel trends assumption.



(a) Grad school enrollment **(b)** High school diploma/GED **(c)** Bachelor's degree

(d) Labor force participation **(e)** Weekly working hours ≥ 35 **(f)** Income (thousands)

Figure 5: Long-run effects of folic acid fortification on young adults' outcomes

Notes: These figures plot event-study estimates with 95% confidence intervals. Figure annotations present the overall effect (point estimate, standard error, and the ratio of the point estimate to the dependent variable's mean). Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). Regressions are weighted by IPUMS person weights. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

All regression estimates are intent-to-treat effects (ITTs). Interpreting the treated group as individuals who avoid CNSA because of fortification, I convert ITTs to treatment effects on the treated (TOTs) by scaling the reduced-form estimates by the fortification-induced

change in the CNSA rate. Figure 4d shows that the CNSA rate falls by 0.20 per 1,000 births in high-baseline-CNSA states and by 0.07 per 1,000 births in low-baseline-CNSA states, implying a first-stage difference of $0.20 - 0.07 = 0.13$. The implied TOTs are therefore $0.008/0.13 = 0.062$ for graduate school enrollment, $-0.007/0.13 = -0.054$ for labor force participation, and $-0.012/0.13 = -0.092$ for income.

Figures A1a-A1f present results for other labor market outcomes. The decline in employment is very close in magnitude to the decline in labor force participation (Figure A1a). Among those who remain in the labor force, however, there is no clear decline in employment (Figure A1b). This suggests that the young adults who move into graduate school are not primarily individuals at the margin of unemployment. Rather, they appear to be individuals who would likely have been employed otherwise. The results for weekly hours are noisier, but they tell a similar story. The decline in hours worked appears to reflect increased graduate school enrollment, not a deterioration in employment prospects among those who would otherwise have been unemployed (Figures 5e and A1d-A1e). Income declines both in the full sample and among employed individuals (Figures 5f and A1f). If the results were driven by weaker local labor markets, we would expect a broader downgrading of labor market outcomes: stronger workers moving into lower-paid jobs and weaker workers being pushed toward unemployment. Section 5.6 tests whether labor market conditions are driving the main results. The evidence does not support the job market explanation. Therefore, these results suggest that the individuals shifting into graduate school would otherwise have been doing reasonably well in the labor market.

Table A2 presents the long-run effects of fortification on several other outcomes, including whether a young adult has a STEM degree, is married, has children if female, lives in a household with income below the poverty threshold, receives food stamp benefits, or works in the state of birth. Overall, I find limited evidence that in utero exposure meaningfully changes these outcomes.

5.3 Treatment dose

Policymakers may also want to know which reductions in folate-related birth defects generate larger effects. To probe this question, this section presents three complementary analyses: (1) comparisons across quartiles of baseline CNSA rates, (2) TWFE dose-response estimates, and (3) dose-specific effects following Callaway, Goodman-Bacon and Sant'Anna (2025).

Comparisons across dose groups. Figures A2a-A2f present event-study estimates comparing states in the fourth, third, and second quartiles of baseline CNSA rates with states in the first quartile, as suggested in Callaway, Goodman-Bacon and Sant'Anna (2024*b*). The shift from work to schooling is sharpest for the Q4-versus-Q1 comparison. Estimates for the other comparisons are much noisier, especially for Q2 relative to Q1. This is consistent with larger effects among young adults born in states with the highest fortification exposure, and further supports using the top quartile of baseline CNSA rates as exposure measure in baseline analysis.

TWFE dose-response estimates. Figures A3a-A3f present the baseline results alongside the dose-response estimates. The dose-response estimates are broadly consistent with the baseline results, though slightly less precise. Given the cautions discussed in Section 4.2, I do not interpret these conventional dose-response estimates as primary evidence. Instead, I treat them as complementary results showing that using a continuous treatment measure does not substantially change the main findings.

Testing SPT assumption. To assess the plausibility of the SPT assumption required for the dose-response estimands, I implement a placebo replication test following the logic of Callaway, Goodman-Bacon and Sant'Anna (2025). For each state in the sample, I construct a placebo-treated unit by duplicating that state's observations and assigning it a unique identifier, while the original sample, including the state itself, serves as the control group. I then estimate a separate event study for each of the approximately 50 placebo replications, using the same specification as our main analysis. Under STP, the pre-period placebo coefficients should be indistinguishable from zero regardless of which state is

designated as treated. Figures A4a-5f overlay the event-study paths for all placebo replications for all six outcomes. The figures show many parallel trend violations, suggesting that the SPT assumption is unlikely to hold in this setting.

5.4 Gender, race, and ethnicity

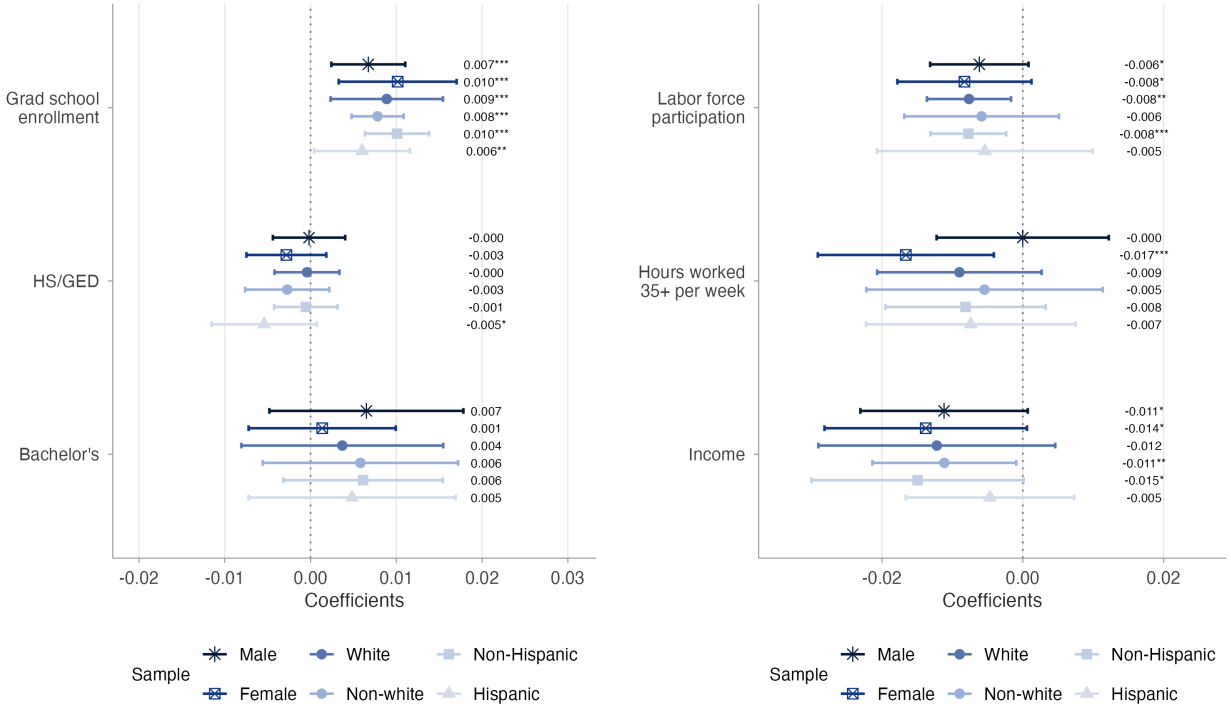
Figures 6a–6b summarize heterogeneity by gender, race (white versus non-white), and ethnicity (Hispanic versus non-Hispanic). Graduate school enrollment rises across all groups, with larger gains among women, white individuals, and non-Hispanic individuals (see also the dynamic effects in Figures A5a–A6c). High school/GED completion and bachelor's degree attainment show little change across any subgroup.

The labor market results mirror the schooling results. Declines in labor force participation and income are larger among women, white individuals, and non-Hispanic individuals (see also the dynamic effects in Figures A6a–A6c and A6g–A6i). An interesting gender difference appears in full-time work: the likelihood of working more than 35 hours per week declines among women but not among men (Figure A6d). This suggests that the shift from work to schooling is concentrated among women who otherwise would have held full-time jobs, whereas the same is not true for men.

Among all subgroups, the effects on graduate school enrollment, labor force participation, and income are smallest for Hispanic individuals. This is consistent with the fact that the 1996 mandate did not cover staple foods widely consumed in Hispanic households, such as corn masa flour and tortillas (Redpath, Kancherla and Oakley, 2018).[8]

[8]The FDA did not permit voluntary fortification of corn masa flour until 2016 (Food and Drug Administration, 2016).

Figure 6: Heterogeneity by gender, race, and ethnicity



(a) Educational outcomes

(b) Labor outcomes

Notes: This table presents cohort-DiD estimates and their 95% confidence interval by gender, eace, and ethnicity. I control for state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. Regressions and dependent-variable means are weighted by IPUMS person weights.

5.5 Robustness

This section shows that the main results are robust to alternative model specifications, alternative thresholds, and baseline periods used to construct fortification exposure, and sample selection. The randomization test and the sensitivity analysis following (Rambachan and Roth, 2023) further suggest that the three precisely estimated effects, on graduate school enrollment, labor force participation, and income, are unlikely to be driven by chance and remain statistically significant under plausible hypothetical violations of parallel trends.

Model specifications. I begin by assessing robustness to alternative model specifications. Online Appendix Figures B1a and B1f show that the results are robust across the

following specifications: (i) a parsimonious model that includes only state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects; (ii) the baseline specification but without controls for confounding policies; (iii) the baseline specification, but with fixed effects entered separately as state-of-birth, survey-year, and quarter-by-year-of-birth fixed effects; and (iv) the baseline specification but additionally controlling for a Bartik-style state-by-year unemployment rate to capture economic conditions at birth.[9]

Exposure thresholds. To assess sensitivity to the exposure definition, I redefine high-baseline-CNSA states using the top 30% and top 20% cutoffs in turn, holding all other specification choices fixed. Online Appendix Figures B2a and B2f show that the main results are largely unchanged across these alternative thresholds.

Baseline periods. The baseline specification uses data from January 1989 through June 1993 to ensure that baseline CNSA rates are measured before the birth of both exposed and unexposed cohorts. As a robustness check, I extend the window used to construct baseline CNSA rates from the birth certificate data to cover 1989 through 1995, the full pre-fortification period. Online Appendix Figures B3a-B3f show that the top-quartile classification is identical across the two windows, and that the results using continuous treatment change only slightly.

Sample selection. Online Appendix Figures B4a and B4f show that the main results are robust to alternative sample choices. I reestimate the models using: (i) PUMS 2015–2024; (ii) PUMS 2019–2024 excluding 2020; and (iii) PUMS 2019–2024 with a wider age range of 22 to 32. For (i), the 2015–2018 PUMS contributes only unexposed cohorts within the age window, which increases the number of pre-treatment cohorts. For (ii), ACS 2020 has a substantially lower response rate than other years, and response rates remain below pre-pandemic levels but are relatively stable through 2024.[10] For (iii), I test whether the results are sensitive to the age window by including younger individuals who may not yet have entered graduate school and older individuals at a different life stage.

[9]The construction of this Bartik-style unemployment rate follows East and Velásquez (2024).

[10]ACS response rates are 94.7% (2016), 93.7% (2017), 92.0% (2018), 86.0% (2019), 71.2% (2020), 85.3% (2021), 84.4% (2022), 84.7% (2023), and 82.9% (2024).

Randomization test. To assess robustness to random noise, I estimate pseudo-effects by randomly reassigning fortification exposure 1,000 times while preserving its empirical distribution across states. Online Appendix Figures B5a and B5f show that the precisely estimated main results lie far in the tails of the pseudo-effect distributions, suggesting that they are unlikely to be driven by chance.

Sensitivity to parallel trend violation. Finally, I assess the sensitivity of the main results to potential violations of the parallel trends assumption using the approach in Rambachan and Roth (2023). For the first exposed cohort, I estimate breakdown values, defined as the magnitude of pre-trend deviations that would render the estimates statistically insignificant at the 90% confidence level. Among the three precisely estimated effects in Figures 5a-5f, the breakdown values are approximately 0.6 for graduate school enrollment and 0.5 for both labor force participation and income (Online Appendix Figures B6a and B6f). Although deviations of this size cannot be ruled out entirely, they appear unlikely given the observed pre-treatment dynamics. Overall, the estimates appear reasonably robust to plausible departures from parallel trends between individuals born in high- and low-baseline-CNSA states.

5.6 Potential mechanisms

This section examines four potential mechanisms and presents suggestive evidence that improved math-related cognitive ability, rather than fertility selection, functional difficulty, or changes in labor market conditions, is most consistent with the shift towards schooling.

Fertility selection. Because fortification reduces birth defects, it may also change the composition of live births, which could contribute to the shift toward schooling. The shift could reflect healthier infants who gain an early advantage in human capital accumulation, or a shift toward births to more advantaged mothers who place greater value on education. To examine this channel, I use birth certificate data to estimate the effects of fortification on birth outcomes and maternal characteristics. Birth outcomes include birth weight, an indicator for low birth weight (birth weight $<$ 2,500 grams), gestational weeks,

and an indicator for preterm birth (gestational weeks < 37). Maternal characteristics include indicators for being non-white, being Hispanic, age 22 or younger, having less than a college education, being unmarried, and having inadequate prenatal care.

Table A3 shows that no improvement birth weight and gestation length among exposed births. Because birth certificate data cover only live births, these results do not imply that fortification has no effect on infant health. Table A4 instead shows an increase in the share of mothers aged 22 or younger, consistent with a shift toward more disadvantaged mothers among live births. If fortification increases survival for infants born to disadvantaged mothers, average birth outcomes among live births can remain unchanged even as outcomes improve for births that would have occurred regardless. Overall, these results suggest that the main findings are unlikely to be driven by healthier infants or by a shift toward more advantaged mothers. If anything, the compositional change could attenuate the estimated effects on later outcomes, since young adults are more likely to have a younger mother.

Functional difficulty. Could fortification have made young adults more likely to pursue advanced degrees by reducing functional difficulty? Using disability variables in PUMS data, I do not find evidence supporting this possibility. Table A7 presents the long-run effects of fortification exposure on disability outcomes and shows little evidence of a decline in disability. The only statistically detectable effect is an increase in reported cognitive disability. However, Figure A7 suggests that this increase is driven entirely by the 1999 cohort. It therefore cannot explain the main effects, which appear across cohorts exposed since 1996. These findings suggest that the shift from work to schooling is unlikely to be driven by changes in disability.

Local labor market conditions. Another possibility is that the shift reflects labor market conditions that coincided with fortification exposure. If exposed cohorts entered adulthood during weaker job markets, they may have been more likely to remain in school, including graduate school, to delay labor market entry when returns to work were temporarily low. To assess this channel, I reestimate the main results while adding

state- or PUMA-of-residence-by-survey-year fixed effects,[11] which absorb time-varying differences in local labor market conditions. If local labor markets were driving the results, the estimated effects should shrink once these fixed effects are included. Table A5 shows that the estimates are essentially unchanged, suggesting that local labor market conditions explain little of the shift.

Table A6 points to the same conclusion. If weak job markets were the main driver, we would expect to see higher unemployment among those in the labor force, or lower job quality among those who remain employed. I find neither. There is no effect on unemployment conditional on labor force participation, and there are no clear changes in occupational quality indices capturing earnings, socioeconomic status, prestige, or the education level associated with the occupation. Although Figure A1f shows that income among the employed does decline, this is better explained by positive selection out of the labor market—those who left for graduate school are precisely the individuals who would have fared relatively well in the labor market—than by deteriorating job market conditions, which would instead predict rising unemployment among labor force participants. Overall, the evidence provides little support for local labor market conditions as the mechanism behind the shift toward graduate schooling.

Cognitive ability. Improvements in cognitive ability may increase educational investment (Murnane, Willett and Levy, 1995; Bowles, Gintis and Osborne, 2001; Heckman, Stixrud and Urzua, 2006). The medical literature suggests that folic acid exposure can improve cognitive development (see Irvine et al. (2022) for a review). To test this channel, I use state-level test score data from the National Assessment of Educational Progress (NAEP) to provide suggestive evidence.[12] Consistent with this mechanism, higher fortification exposure is associated with higher math scores among 4th and 8th graders.

I begin by mapping fortification exposure from state of birth to state of school, since

[11] PUMA: Public Use Microdata Area, the smallest geographic unit in the ACS PUMS. In the sample, over 95% of respondents work in their state of residence.

[12] Restricted-use NAEP microdata are ordinarily accessible by application through the Institute of Education Sciences at the U.S. Department of Education, but access is currently paused. I applied for restricted-use student-level data, but on April 2, 2025, I was informed that all applications were paused. I do not expect access to resume soon.

the public NAEP data report only state of school, survey year, and grade. I construct a migration-adjusted exposure measure that weights birth-state exposure by the observed composition of student birth places within each state-grade-year cell using ACS PUMS from the same survey years as the NAEP files. Let j index state of school, $g \in \{4, 8\}$ grade, c survey year, and s state of birth (with all foreign countries grouped together). Define $\pi_{s|j,g,c}$ as the share of students in (j, g, c) who were born in s, and let $\mathbf{1}\{\text{High CNSA}\}_s$ indicate whether birth state s is in the top quartile of baseline CNSA rates. The exposure used for NAEP cell (j, g, c) is

$$P(\text{High CNSA})_{j,g,c} = \sum_s \pi_{s|j,g,c} \times \mathbf{1}\{\text{High CNSA}\}_s,$$

with the indicator set to 0 for foreign born. This yields a state of school exposure measure that accounts for observed migration patterns among 4th and 8th graders, and captures the likelihood that students in state j were born in a high-baseline-CNSA state.

I assign representative ages of 10 (grade 4) and 14 (grade 8) to back out likely birth cohorts and define exposed cohorts as those born after 1996. Online Appendix Figures B7a and B7b show that roughly 90.8% of 4th graders are ages 9 to 10 and roughly 89.3% of 8th graders are ages 13 to 14. This assignment likely biases exposure downward, so the estimates are conservative. The results are similar when I instead assign ages of 9 (grade 4) and 13 (grade 8) (Online Appendix Table B1).

The regression model is:

$$Y_{jgc} = \zeta\big(P(\text{High CNSA})_{jgc} \times \text{Exposed cohorts}_{gc}\big) + \eta_{jc} + \theta_g + Z_{jgc} + \varepsilon_{jgc}, \qquad (2)$$

where Y_{jgc} is the average test score outcome for state j in survey year c and grade g; $P(\text{High CNSA})_{jgc}$ is the migration-adjusted exposure; and $\text{Exposed cohorts}_{gc}$ is an indicator for likely birth cohorts after 1996.[13] η_{jc} includes state-of-school-by-likely-year-of-birth fixed effects that absorb unobserved factors common to students who attend school

[13] Because quarter of birth is not observed in NAEP, I define exposed cohorts as students likely born after 1996.

in the same state and belong to the same cohort, including time-varying differences in school policies, resources, and local economic conditions. θ_g are grade fixed effects. Z_{jgc} includes migration-adjusted exposure measures for potentially confounding policies, including Medicaid and CHIP expansions, welfare reform, and state mental health parity laws, constructed using the same weights as in $P(\text{High CNSA})_{jgc}$. I pool 4th and 8th graders and focus on a pre-versus-post comparison because the aggregate data include a limited number of state-by-grade-by-year cells.

Table 2 reports the estimated relationship between in utero exposure to folic acid fortification and NAEP performance among 4th and 8th graders. Columns (1) to (3) show positive associations between fortification exposure and math scores: when the likelihood of being born in a high-baseline-CNSA state moves from 0 to 1, average math scores rise by 4.51 points (1.73% at the mean), the share scoring at or above NAEP Basic increases by 3.66 percentage points, and the share at or above NAEP Proficient increases by 4.63 percentage points. I find no noticeable changes in reading scores or in the share at or above NAEP Basic. These results suggest that improved math-related cognitive ability is the likely mechanism behind the increase in graduate school enrollment.

Table 2: In utero fortification exposure and test scores of 4th and 8th graders

	Math			Reading		
	Average score	% ≥ Basic	% ≥ Profi-cient	Average score	% ≥ Basic	% ≥ Profi-cient
	(1)	(2)	(3)	(4)	(5)	(6)
P(High CNSA)×Exposed cohorts	4.508***	3.661***	4.632***	-0.822	1.140	-1.009
	(1.130)	(1.104)	(0.902)	(1.095)	(1.478)	(0.610)
Observations	408	408	408	408	408	408
R^2	0.995	0.945	0.938	0.995	0.951	0.945
Dep. var. mean	259.991	75.502	34.361	241.256	69.351	31.305
Est./Dep. var. mean	1.734%	4.849%	13.480%	-0.341%	1.644%	-3.224%
State-of-school-by-(likely)-year-of-birth FEs	✓	✓	✓	✓	✓	✓
Grade FEs	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓

Notes: This table presents cohort-DiD estimates with standard errors clustered at state of school. $P(\text{High CNSA})$ is fortification exposure adjusted for migration. Exposed cohorts are those with likely year of birth after 1996, determined by assigning age 10 to grade 4 and age 14 to grade 8. Controls include state-of-school-by-(likely)-year-of-birth, grade fixed effects, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. The unit of observation is state-by-year-by-grade cells. Regressions are weighted by student counts from ACS PUMS. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

6 Marginal Value of Public Funds

I follow Hendren and Sprung-Keyser (2020) to calculate the marginal value of public funds (MVPF) for folic acid fortification and comparison nutrition policies. The MVPF is defined as the ratio of beneficiaries' willingness to pay (WTP) for a policy to the policy's net cost to the government: $MVPF_p = \frac{WTP_p}{C_p - \tau WTP_p}$, where C_p is the direct implementation cost of policy p, τ is the tax-and-transfer rate, and τWTP_p captures the fiscal feedback from higher earnings. This framework measures how much welfare a policy delivers to beneficiaries per dollar of net public spending and provides a common metric for comparing policies. I use a tax-and-transfer rate of 18.9% following Hendren and Sprung-Keyser (2020), a discount rate of 3% following Office of Management and Budget (2023), and baseline annual earnings of $35,000. All costs and benefits are expressed in 2000 dollars.

For each policy, I measure WTP as the present value of the long-run earnings gain attributable to exposure. I approximate the annual dollar gain as $\Delta Y_p \simeq \bar{Y}\hat{\beta}_p$, where $\bar{Y}$ is

baseline annual earnings and $\hat{\beta}_p$ is the estimated effect on log adult earnings or income. I then treat this annual gain as a 30-year annuity from ages 30 to 60 and discount it back to birth: $WTP_p = \sum_{a=30}^{59} \frac{\bar{Y}\hat{\beta}_p}{(1+r)^a} = \bar{Y}\hat{\beta}_p \sum_{a=30}^{59}(1+r)^{-a}$. Using $r = 0.03$, the present-value annuity factor is $\sum_{a=30}^{59}(1.03)^{-a} = 8.317$. Thus, $WTP_p \simeq 8.317 \times \bar{Y}\hat{\beta}_p$. The implied fiscal feedback from higher adult earnings is $TaxRevenue_p = \tau WTP_p$. This calculation is conservative because it excludes earnings gains before age 30 and after age 60, as well as other potential benefits such as improved health, lower special education costs, and reduced medical spending.

To place the magnitude of folic acid fortification in context, I compare its implied earnings effect with estimates from other large-scale nutrition policies. For folic acid fortification, I convert the estimated effect on graduate school enrollment into an implied effect on earnings using the estimated return to a graduate degree from Altonji, Arcidiacono and Maurel (2016) (0.159). The implied effect on log earnings is $\hat{\beta}_{FA} = \frac{0.008}{0.13} \times 0.159 \approx 0.010$. With baseline annual earnings of \$35,000, this corresponds to an annual earnings gain of about $\$35{,}000 \times 0.010 = \$350$. Discounting a 30-year stream of these gains from ages 30 to 60 back to birth gives $WTP_{FA} = \$350 \times 8.317 \approx \$2{,}911$. The implied fiscal feedback from higher adult earnings is $TaxRevenue_{FA} = 0.189 \times \$2{,}911 \approx \$550$.

For comparison, Adhvaryu et al. (2020) estimate that in utero exposure to salt iodization increased adult income by 10.5% for fully exposed cohorts and by 2.7% for partially exposed cohorts. I use the average of these two estimates, 6.6%, to approximate the income gain among exposed cohorts. Applying the same annuity calculation implies a present value of earnings gains of $8.317 \times \$35{,}000 \times 0.066 \approx \$19{,}211$. Niemesh (2015) estimates that in utero exposure to iron fortification increased adult income by 1.4%, implying a present value of earnings gains of $8.317 \times \$35{,}000 \times 0.014 \approx \$4{,}075$. Bailey et al. (2024) estimates that exposure to Food Stamps from conception through age five increased adult income by 7.1%, implying a present value of earnings gains of $8.317 \times \$35{,}000 \times 0.071 \approx \$20{,}665$. I summarize these estimates in Figure 7. These estimates are broadly in the same order of magnitude, although they are not directly comparable because the policies differ in target populations, exposure definitions, ages at

outcome measurement, and policy margins.

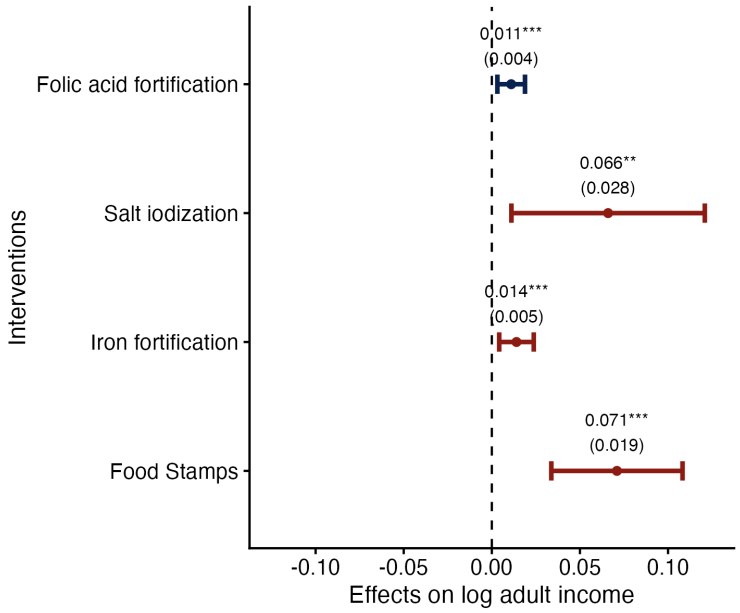


Figure 7: Long-run effects on log adult income by nutrition policy

Notes: This figure plots point estimates for each nutrition intervention's effect on log adult income, with 95% confidence intervals. The folic acid estimate is constructed by multiplying the estimated effect on graduate school enrollment by the estimated return to graduate degrees from Altonji, Arcidiacono and Maurel (2016); all other values are regression coefficients from the cited studies. Adhvaryu et al. (2020) use baseline goiter prevalence as a continuous proxy for iodine deficiency, while Niemesh (2015) use estimated iron consumption as a proxy for iron deficiency. Bailey et al. (2024) reports the effect of Food Stamps exposure from conception through age five.

The fiscal cost of these fortification policies is very small relative to their implied long-run earnings benefits. For salt iodization, the historical cost in the U.S. was borne privately by the Morton Salt Company, but I include a cost estimate to approximate the MVPF of a government-financed iodization policy. World Health Organization (2014) estimates the cost of salt iodization at \$0.02–\$0.05 per person covered in 2014 dollars; I use the upper bound of \$0.05. For iron fortification, Wilder and Williams (1944) reports that enrichment cost 35–50 cents per person annually in 1943. For folic acid fortification, Grosse et al. (2005) estimates direct fortification costs of about \$3 million per year.

Because these direct costs are small, the implied MVPFs are very large. After discounting the 30-year earnings stream back to birth, the fiscal feedback from higher earnings exceeds the direct implementation cost for these fortification policies. Under the MVPF

convention, policies with positive WTP and nonpositive net fiscal cost are treated as having an effectively infinite MVPF, since they generate welfare gains without requiring net government spending. Compared with broader nutrition programs such as Food Stamps, whose MVPF is estimated to be 1.04 in Hendren and Sprung-Keyser (2020) and 62.25 in Bailey et al. (2024), fortification policies appear especially cost-effective because they are far cheaper to implement.

7 Conclusion

This paper provides the first evidence on the long-run human capital effects of reducing folate deficiency, which operates in the earliest stages of fetal development, by combining cross-state variation in baseline rates of folate-linked birth defects with variation in first-trimester timing relative to the fortification mandate. Results show a shift from work toward schooling among exposed young adults: among those aged 23–30, graduate school enrollment increases by 0.8 percentage points, while labor force participation falls by 0.7 percentage points and annual income decreases by $12. Effects are larger among women, White individuals, and non-Hispanic individuals. Beyond being less likely to participate in the labor force, exposed women are also less likely to work full time. Estimates for Hispanic individuals are smaller, consistent with weaker exposure because the 1996 mandate did not cover staple foods widely consumed in Hispanic households. Results on fertility selection suggest that fortification is associated with a small increase in births to mothers younger than 22, which may attenuate the estimated effects. Evidence on mechanisms points away from functional difficulty or changes in local labor market conditions as potential explanations. Instead, fortification exposure is associated with higher math scores among 4th- and 8th-grade students, consistent with improved math-related cognitive ability as a plausible driver of the shift toward schooling. Back-of-the-envelope calculations suggest that the implied adult income gain from folic acid fortification is smaller than, but broadly comparable in magnitude to, those from salt iodization, iron fortification, and Food Stamps. The additional annual tax revenue from adult earnings gains alone can cover fortification costs, giving folic acid fortification an effectively infinite marginal value of public funds and making it highly fiscally appealing.

This study has several limitations. First, the exposure measure is constructed from live births rather than all pregnancies. Data on the full universe of pregnancies are not available; fetal death files primarily cover fetuses at 20 weeks or later, and many states began reporting congenital anomalies in these records later than in birth certificates. Second, exposed cohorts are still too young to permit study of downstream adult outcomes such as prime-age earnings or family formation. Following these cohorts over the life course would provide a more complete assessment of fortification's effects. Third, because the design compares higher- and lower-exposure groups rather than treated and untreated groups, estimates capture relative rather than absolute effects. Despite these limitations, the results speak to how prenatal nutrition shapes downstream human capital, aligning with the fetal-origins literature in showing that early-life interventions can generate long-run gains.

Compared with large programs such as Food Stamps, which broadly subsidize food consumption, fortification targets specific micronutrient deficiencies at much lower cost yet delivers meaningful benefits, making it far more cost-effective. This contrast highlights potential efficiency gains from more targeted nutrition policy—concentrating resources on nutrients and foods with the highest marginal returns rather than subsidizing healthy foods more broadly—a consideration that is particularly relevant in an era of tightening welfare budgets (Bergh, 2025).

This paper also suggests scope for larger gains from folic acid fortification in low-income countries, where folate deficiency is often more severe (McLean, de Benoist and Allen, 2008). The U.S. experience is encouraging, but more evidence is needed on how fortification performs in other settings, particularly where diets, staple foods, and the feasibility and quality of fortification technologies may differ.

References

Abdel-Salam, Ghada, and Andrew E Czeizel. 2000. "A case-control etiologic study of microcephaly."

Adhvaryu, Achyuta, Steven Bednar, Teresa Molina, Quynh Nguyen, and Anant Nyshadham. 2020. "When it rains it pours: The long-run economic impacts of salt iodization in the United States." *Review of Economics and Statistics*, 102(2): 395–407.

Ahluwalia, Namanjeet, Johanna Dwyer, Ana Terry, Alanna Moshfegh, and Clifford Johnson. 2016. "Update on NHANES dietary data: focus on collection, release, analytical considerations, and uses to inform public policy." *Advances in Nutrition*, 7(1): 121–134.

Almond, Douglas, and Bhashkar Mazumder. 2011. "Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy." *American Economic Journal: Applied Economics*, 3(4): 56–85.

Almond, Douglas, Bhashkar Mazumder, and Reyn Van Ewijk. 2015. "In utero Ramadan exposure and children's academic performance." *The Economic Journal*, 125(589): 1501–1533.

Almond, Douglas, Lena Edlund, Hongbin Li, and Junsen Zhang. 2007. "Long-term effects of the 1959-1961 China famine: Mainland China and Hong Kong."

Altonji, Joseph G., Peter Arcidiacono, and Arnaud Maurel. 2016. "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects." In *Handbook of the Economics of Education*. Vol. 5, , ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 305–396. Elsevier.

American College of Obstetricians and Gynecologists. 2017. "Practice Bulletin No. 187: Neural Tube Defects." *Obstetrics & Gynecology*, 130(6): e279–e290.

Anderson, Ellen, Betty Perloff, Jaspreet KC Ahuja, and Nancy Raper. 2001. "Tracking nutrient changes for trends analysis in the United States." *Journal of Food Composition and Analysis*, 14(3): 287–294.

Araújo, Daniel, Bladimir Carrillo, and Breno Sampaio. 2021. "The long-run economic consequences of iodine supplementation." *Journal of Health Economics*, 79: 102490.

Atwood, Alicia. 2022. "The long-term effects of measles vaccination on earnings and employment." *American Economic Journal: Economic Policy*, 14(2): 34–60.

Bailey, Martha J, Hilary Hoynes, Maya Rossin-Slater, and Reed Walker. 2024. "Is the social safety net a long-term investment? Large-scale evidence from the food stamps program." *Review of Economic Studies*, 91(3): 1291–1330.

Bedard, Kelly, and Douglas A Herman. 2008. "Who goes to graduate/professional school? The importance of economic fluctuations, undergraduate field, and ability." *Economics of Education Review*, 27(2): 197–210.

Bentley, Tanya GK, Milton C Weinstein, Walter C Willett, and Karen M Kuntz. 2009. "A cost-effectiveness analysis of folic acid fortification policy in the United States." *Public Health Nutrition*, 12(4): 455–467.

Bergh, Katie. 2025. "Millions of Low-Income Households Would Lose Food Aid Under Proposed House Republican SNAP Cuts." *Washington, DC: Center on Budget and Policy Priorities*.

Berry, RJ, J Mulinare, and HC Hamner. 2010. "Folic acid fortification: neural tube defect risk reduction—a global perspective." *Folate in Health and Disease*, 2.

Black, Dan A, Terra G McKinnish, and Seth G Sanders. 2005. "Tight labor markets and the demand for education: Evidence from the coal boom and bust." *ILR Review*, 59(1): 3–16.

Black, Sandra E, Lesley J Turner, and Jeffrey T Denning. 2023. "PLUS or minus? The effect of graduate school loans on access, attainment, and prices." National Bureau of Economic Research.

Bleakley, Hoyt. 2007. "Disease and development: evidence from hookworm eradication in the American South." *Quarterly Journal of Economics*, 122(1): 73–117.

Bleakley, Hoyt. 2010. "Malaria eradication in the Americas: A retrospective analysis of childhood exposure." *American Economic Journal: Applied Economics*, 2(2): 1–45.

Blumenfeld, Zeev, Ephraim Siegler, and Moshe Bronshtein. 1993. "The early diagnosis of neural tube defects." *Prenatal Dagnosis*, 13(9): 863–871.

Bowles, Samuel, Herbert Gintis, and Melissa Osborne. 2001. "The determinants of earnings: A behavioral approach." *Journal of Economic Literature*, 39(4): 1137–1176.

Bureau of Economic Analysis. 1988. "Regional Economic Information System (REIS): Historical Data: Local Area Personal Income and Employment: Personal Current Transfer Receipts, 1988."

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2024*a*. "Difference-in-differences with a continuous treatment." National Bureau of Economic Research.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2024*b*. "Event studies with a continuous treatment." Vol. 114, 601–605, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro Sant'Anna. 2025. "Difference-in-Differences with a Continuous Treatment." arXiv. org.

Charles, Kerwin Kofi, Erik Hurst, and Matthew J. Notowidigdo. 2018. "Housing Booms and Busts, Labor Market Opportunities, and College Attendance." *American Economic Review*, 108(10): 2947–2994.

Chen, Yuyu, and Li-An Zhou. 2007. "The long-term health and economic consequences of the 1959–1961 famine in China." *Journal of Health Economics*, 26(4): 659–681.

Committee on Use of Dietary Reference Intakes in Nutrition Labeling. 2004. *Dietary reference intakes: guiding principles for nutrition labeling and fortification.* National Academies Press.

Cragan, Janet D, Helen E Roberts, Larry D Edmonds, Muin J Khoury, Russell S Kirby, Gary M Shaw, Ellen M Velie, Ruth D Merz, Mathias B Forrester, Roger A Williamson, et al. 1995. "Surveillance for anencephaly and spina bifida and the impact of prenatal diagnosis–United States, 1985-1994." *MMWR. CDC Surveillance Summaries: Morbidity and Mortality Weekly Report. CDC Surveillance Summaries*, 44(4): 1–13.

Czeizel, Andrew E. 2000. "Primary prevention of neural-tube defects and some other major congenital abnormalities: recommendations for the appropriate use of folic acid during pregnancy." *Paediatric Drugs*, 2: 437–449.

Deng, Zichen, and Maarten Lindeboom. 2022*a*. "A bit of salt, a trace of life: Gender norms and the impact of a salt iodization program on human capital formation of school aged children." *Journal of Health Economics*, 83: 102614.

Deng, Zichen, and Maarten Lindeboom. 2022*b*. "Early-life famine exposure, hunger recall, and later-life health." *Journal of Applied Econometrics*, 37(4): 771–787.

East, Chloe N. 2020. "The effect of food stamps on children's health evidence from immigrants' changing eligibility." *Journal of Human Resources*, 55(2): 387–427.

East, Chloe N, and Andrea Velásquez. 2024. "Unintended consequences of immigration enforcement: household services and high-educated mothers' work." *Journal of Human Resources*, 59(5): 1458–1502.

Feyrer, James, Dimitra Politi, and David N Weil. 2017. "The cognitive effects of micronutrient deficiency: Evidence from salt iodization in the United States." *Journal of the European Economic Association*, 15(2): 355–387.

Field, Erica, Omar Robles, and Maximo Torero. 2009. "Iodine deficiency and schooling attainment in Tanzania." *American Economic Journal: Applied Economics*, 1(4): 140–69.

Finer, Lawrence B, and Mia R Zolna. 2016. "Declines in unintended pregnancy in the United States, 2008–2011." *New England Journal of Medicine*, 374(9): 843–852.

Fitzsimons, Emla, and Marcos Vera-Hernández. 2022. "Breastfeeding and child development." *American Economic Journal: Applied Economics*, 14(3): 329–66.

Food and Drug Administration. 1996. "Food standards: Amendment of standards of identity for enriched grain products to require addition of folic acid; final rule (21 CFR Parts 136, 137, and 139)." *Federal Register*, 61: 8781–8797.

Food and Drug Administration. 2015. "Questions and answers on FDA's fortification policy: guidance for industry."

Food and Drug Administration. 2016. "Food additives permitted for direct addition to food for human consumption; folic acid. Final rule." *Federal register*, 81(73): 22176–22183.

Global Fortification Data Exchange. 2025. "Map: Countries with mandatory or voluntary legislation status." Accessed April 25, 2026.

Greene, Nicholas DE, and Andrew J Copp. 2014. "Neural tube defects." *Annual Review of Neuroscience*, 37: 221–242.

Grosse, Scott D, Norman J Waitzman, Patrick S Romano, and Joseph Mulinare. 2005. "Reevaluating the benefits of folic acid fortification in the United States: economic analysis, regulation, and public health." *American Journal of Public Health*, 95(11): 1917–1922.

Heckman, James J, Jora Stixrud, and Sergio Urzua. 2006. "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics*, 24(3): 411–482.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. "A unified welfare analysis of government policies." *The Quarterly Journal of Economics*, 135(3): 1209–1318.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2016. "Long-run impacts of childhood access to the safety net." *American Economic Review*, 106(4): 903–934.

Hoynes, Hilary, Marianne Page, and Ann Huff Stevens. 2011. "Can targeted transfers improve birth outcomes?: Evidence from the introduction of the WIC program." *Journal of Public Economics*, 95(7-8): 813–827.

Huang, Qingyang, Chang Liu, and Li-An Zhou. 2020. "Farewell to the God of Plague: Estimating the effects of China's Universal Salt Iodization on educational outcomes." *Journal of Comparative Economics*, 48(1): 20–36.

Hutt, Peter Barton. 1984. "Government regulation of the integrity of the food supply." *Annual Review of Nutrition*, 4(1): 1–21.

International Monetary Fund. 2026. "Fiscal Monitor: Fiscal Policy under Pressure: High Debt, Rising Risks."

Irvine, Nathalie, Gillian England-Mason, Catherine J Field, Deborah Dewey, and Fariba Aghajafari. 2022. "Prenatal folate and choline levels and brain and cognitive development in children: a critical narrative review." *Nutrients*, 14(2): 364.

Jacques, Paul F, Jacob Selhub, Andrew G Bostom, Peter WF Wilson, and Irwin H Rosenberg. 1999. "The effect of folic acid fortification on plasma folate and total homocysteine concentrations." *New England Journal of Medicine*, 340(19): 1449–1454.

Johnson, Matthew T. 2013. "The impact of business cycle fluctuations on graduate school enrollment." *Economics of Education Review*, 34: 122–134.

Kancherla, Vijaya, Lorenzo D Botto, Laura A Rowe, Nathan A Shlobin, Adrian Caceres, Anastasia Arynchyna-Smith, Kathrin Zimmerman, Jeffrey Blount, Zewdie Kibruyisfaw, Kemel A Ghotme, et al. 2022. "Preventing birth defects, saving lives, and promoting health equity: an urgent call to action for universal mandatory food fortification with folic acid." *The Lancet Global Health*, 10(7): e1053–e1057.

Kuecken, Maria, Josselin Thuilliez, and Marie-Anne Valfort. 2021. "Disease and human capital accumulation: Evidence from the Roll Back Malaria partnership in Africa." *The Economic Journal*, 131(637): 2171–2202.

Lazuka, Volha. 2020. "Infant Health and Later-Life Labor Market Outcomes Evidence from the Introduction of Sulfa Antibiotics in Sweden." *Journal of Human Resources*, 55(2): 660–698.

Lewis, Christine J, Nancy T Crane, Dennis B Wilson, and Elizabeth A Yetley. 1999. "Estimated folate intakes: data updated to reflect food fortification, increased bioavailability, and dietary supplement use." *The American Journal of Clinical Nutrition*, 70(2): 198–207.

Lindeboom, Maarten, France Portrait, and Gerard J Van den Berg. 2010. "Long-run effects on longevity of a nutritional shock early in life: the Dutch Potato famine of 1846–1847." *Journal of Health Economics*, 29(5): 617–629.

Liu, Jufen, Lei Jin, Zhiwen Li, Yali Zhang, Le Zhang, Linlin Wang, and Aiguo Ren. 2018. "Prevalence and trend of isolated and complicated congenital hydrocephalus and preventive effect of folic acid in northern China, 2005–2015." *Metabolic Brain Disease*, 33: 837–842.

Majid, Muhammad Farhan. 2015. "The persistent effects of in utero nutrition shocks over the life cycle: Evidence from Ramadan fasting." *Journal of Development Economics*, 117: 48–57.

McLean, Erin, Bruno de Benoist, and Lindsay H Allen. 2008. "Review of the magnitude of folate and vitamin B12 deficiencies worldwide." *Food and Nutrition Bulletin*, 29(2_suppl1): S38–S51.

McNulty, Helene, and Kristina Pentieva. 2004. "Folate bioavailability." *Proceedings of the Nutrition Society*, 63(4): 529–536.

Mersereau, P, K Kilker, H Carter, E Fassett, J Williams, A Flores, C Prue, L Williams, C Mai, and J Mulinare. 2004. "Spina Bifida and Anencephaly Before and After Folic Acid Mandate–United States, 1995–1996 and 1999–2000." *MMWR: Morbidity & Mortality Weekly Report*, 53(17).

Murnane, Richard J, John B Willett, and Frank Levy. 1995. "The growing importance of cognitive skills in wage determination." *The Review of Economics and Statistics*, 77(2): 251.

National Center for Health Statistics. 2003. "All-county natality files for 1989-2003." as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.

Naz, Naila, Alicia Requena Jimenez, Anna Sanjuan-Vilaplana, Megan Gurney, and Jaleel Miyan. 2016. "Neonatal hydrocephalus is a result of a block in folate handling and metabolism involving 10-formyltetrahydrofolate dehydrogenase." *Journal of Neurochemistry*, 138(4): 610–623.

Nguyen, Vy Kim, Lauren YM Middleton, Lei Huang, Neil Zhao, Eliseu Verly Jr, Jacob Kvasnicka, Luke Sagers, Chirag J Patel, Justin Colacino, and Olivier Jolliet. 2023. "Harmonized US National Health and Nutrition Examination Survey 1988-2018 for high throughput exposome-health discovery." *MedRxiv*.

Niemesh, Gregory T. 2015. "Ironing out deficiencies: evidence from the united states on the economic effects of iron deficiency." *Journal of Human Resources*, 50(4): 910–958.

Obeid, Rima, Wolfgang Holzgreve, and Klaus Pietrzik. 2013. "Is 5-methyltetrahydrofolate an alternative to folic acid for the prevention of neural tube defects?" *Journal of Perinatal Medicine*, 41(5): 469–483.

Office of Management and Budget. 2023. "Circular No. A-94: Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs." Executive Office of the President Circular A-94, Washington, DC.

Petrini, JR, K Damus, and RB Johnston. 1999. "Knowledge and Use of Folic Acid by Women of Childbearing Age–United States, 1995 and 1998." *MMWR: Morbidity & Mortality Weekly Report*, 48(16): 325–327.

Pfeiffer, Christine M, Jeffery P Hughes, Ramon A Durazo-Arvizu, David A Lacher, Christopher T Sempos, Mindy Zhang, Elizabeth A Yetley, and Clifford L Johnson. 2012. "Changes in measurement procedure from a radioassay to a microbiologic assay necessitate adjustment of serum and RBC folate concentrations in the US population from the NHANES 1988–2010." *The Journal of Nutrition*, 142(5): 894–900.

Quinlivan, EP, J McPartlin, H McNulty, M Ward, JJ Strain, DG Weir, and JM Scott. 2002. "Importance of both folic acid and vitamin B12 in reduction of risk of vascular disease." *The Lancet*, 359(9302): 227–228.

Rambachan, Ashesh, and Jonathan Roth. 2023. "A more credible approach to parallel trends." *Review of Economic Studies*, 90(5): 2555–2591.

Ray, Joel G, Gita Singh, and Robert F Burrows. 2004. "Evidence for suboptimal use of periconceptional folic acid supplements globally." *BJOG: An International Journal of Obstetrics & Gynaecology*, 111(5): 399–408.

Redpath, Ben, Vijaya Kancherla, and Godfrey P Oakley. 2018. "Availability of corn masa flour and tortillas fortified with folic acid in Atlanta after national regulations allowing voluntary fortification." *JAMA*, 320(15): 1600–1601.

Rossin-Slater, Maya. 2013. "WIC in your neighborhood: New evidence on the impacts of geographic access to clinics." *Journal of Public Economics*, 102: 51–69.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, Renae Rodgers, Jonathan Schroeder, and Kari C.W. Williams. 2025. "Ipums usa: Version 16.0 [dataset]."

Scholte, Robert S, Gerard J Van Den Berg, and Maarten Lindeboom. 2015. "Long-run effects of gestation during the Dutch Hunger Winter famine on labor market and hospitalization outcomes." *Journal of Health Economics*, 39: 17–30.

Serena, Benjamin Ly. 2019. "Cognitive consequences of iodine deficiency in adolescence: evidence from salt iodization in Denmark." *The Scandinavian Journal of Economics*.

Smithells, RW, MJ Seller, R Harris, DW Fielding, CJ Schorah, NC Nevin, S Sheppard, AP Read, S Walker, and J Wild. 1983. "Further experience of vitamin supplementation for prevention of neural tube defect recurrences." *The Lancet*, 321(8332): 1027–1031.

Tafesse, Wiktoria. 2022. "The effect of Universal Salt Iodization on cognitive test scores in rural India." *World Development*, 152: 105796.

Toivonen, KI, E Lacroix, M Flynn, PE Ronksley, KA Oinonen, A Metcalfe, and TS Campbell. 2018. "Folic acid supplementation during the preconception period: a systematic review and meta-analysis." *Preventive Medicine*, 114: 1–17.

US Bureau of Labor Statistics. 1989-2002. "Quarterly Census of Employment and Wages: 1989-2002."

US Census Bureau. 1990. "Intercensal Population Estimates: 1980-1990."

US Census Bureau. 2009. "County and City Data Book [United States], 1988."

Wald, NJ, MR Law, JK Morris, and DS Wald. 2001. "Quantifying the effect of folic acid." *The Lancet*, 358(9298): 2069–2073.

Weiler, William C. 1994. "Expectations, undergraduate debt and the decision to attend graduate school: A simultaneous model of student choice." *Economics of Education Review*, 13(1): 29–41.

Wilder, RM, and RR Williams. 1944. "Enrichment of flour and bread: a history of the movement." *Bulletin: National research council. National Academy of Sciences*, , (110).

Williams, Jennifer, Cara T Mai, Joe Mulinare, Jennifer Isenburg, Timothy J Flood, Mary Ethen, Barbara Frohnert, Russell S Kirby, Centers for Disease Control, Prevention, et al. 2015. "Updated estimates of neural tube defects prevented by mandatory folic acid fortification—United States, 1995–2011." *MMWR Morb Mortal Wkly Rep*, 64(1): 1–5.

World Health Organization. 2014. "Guideline: Fortification of Food-Grade Salt with Iodine for the Prevention and Control of Iodine Deficiency Disorders." https://apps.who.int/iris/handle/10665/136908 (accessed on 13 May 2022).

Wright, Kevin. 2003. "Folic Acid and the American Food Supply: A historical account of the FDA's creation of the current folic acid regulations."

Yetley, Elizabeth A, Christine M Pfeiffer, Karen W Phinney, Zia Fazili, David A Lacher, Regan L Bailey, Sheena Blackmore, Jay L Bock, Lawrence C Brody, Ralph Carmel, et al. 2011. "Biomarkers of folate status in NHANES: a roundtable summary." *The American Journal of Clinical Nutrition*, 94(1): 303S–312S.

Yi, Yunni, Marion Lindemann, Antje Colligs, and Claire Snowball. 2011. "Economic burden of neural tube defects and impact of prevention with folic acid: a literature review." *European Journal of Pediatrics*, 170(11): 1391–1400.

Appendix

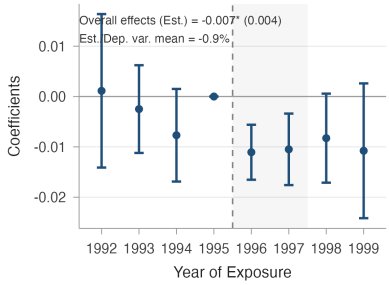
A Tables and figures

Table A1: Fortification exposure and pre-1989 state/county characteristics

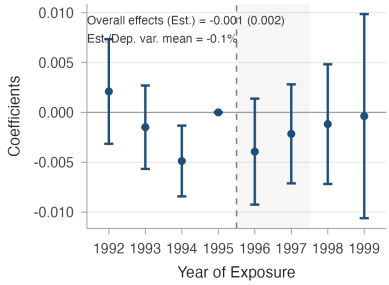
	High $CNSA_s$		
	State-level	County level	
	(1)	(2)	(3)
Demographic features			
Share of black (%), 1988	-0.002	-0.002	0.000
	(0.015)	(0.002)	(0.002)
Share of female (%), 1988	0.015	0.037*	0.008
	(0.204)	(0.021)	(0.019)
Share of under 5 (%), 1988	0.260	0.048*	-0.030
	(0.217)	(0.025)	(0.020)
Share of over 65 (%), 1988	-0.006	-0.011	-0.016
	(0.081)	(0.012)	(0.012)
Birth rate (%), 1988	-0.022	0.012	0.014**
	(0.057)	(0.008)	(0.006)
Death rate (%), 1988	0.162	0.008	-0.013
	(0.272)	(0.020)	(0.020)
Log population, 1988	0.093	-0.031	0.005
	(0.124)	(0.026)	(0.021)
Economic conditions			
Transfer income p.p. (million $), 1988	-0.307	-0.007	0.129
	(0.666)	(0.147)	(0.091)
Income p.p. (million $), 1985	107.878	23.606	3.413
	(113.606)	(17.774)	(18.994)
Federal funds p.p. (million $), 1986	-234.162	9.907	8.859
	(162.300)	(9.340)	(9.560)
Unemployment rate (%), 1986	-0.056	-0.005	-0.005
	(0.066)	(0.011)	(0.006)
Agriculture			
Value of produces sold per farm (million $), 1987	-4.560	-0.596*	-0.028
	(2.789)	(0.322)	(0.129)
Average farm size (million acres), 1987	152.637	-13.204	-0.314
	(121.444)	(8.099)	(4.841)
State FEs			Y
Observations	49	2,850	2,850
R^2	0.321	0.051	0.283
Dep. var. mean	0.294	0.224	0.224

Notes: This table presents results from regressing high CNSA—an indicator for states in the 4th quartile of baseline CNSA rates—on pre-1989 state or county characteristics. Regressions are weighted by 1988 population. Data sources: County Intercensal Estimates (demographic shares and population), County Databook 1988 (birth rate, death rate, farm value, and farm size), BEA Regional Economic Information System (transfers), and Bureau of Labor Statistics (unemployment). Standard errors in county-level regressions are clustered at the state level. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

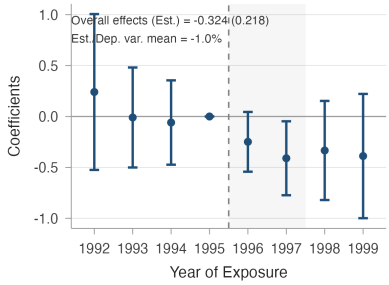
Figure A1: Additional labor outcomes



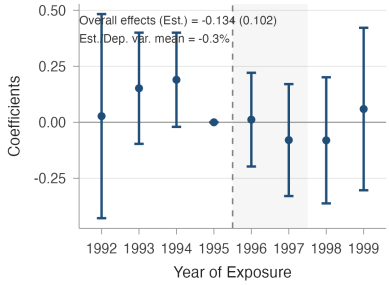
(a) Employed, full sample



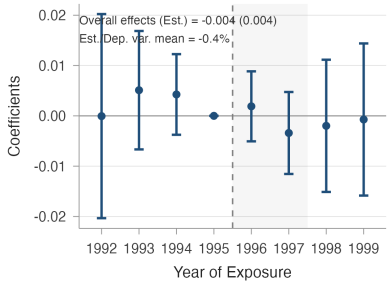
(b) Employed among those in labor force



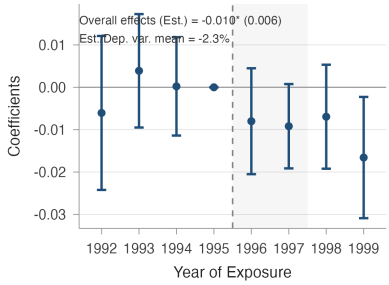
(c) Weekly working hours, full sample



(d) Weekly working hours among those employed



(e) Weekly working hours ≥ 35 among those employed



(f) Income (thousands) among those employed

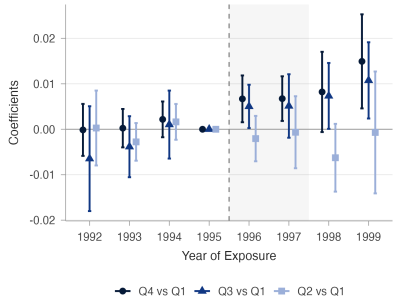
Notes: These figures plot event-study estimates with 95% confidence intervals. Figure annotations present the overall effect (point estimate, standard error, and the ratio of the point estimate to the dependent variable's mean). Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). Regressions are weighted by IPUMS person weights. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

Table A2: Other outcomes

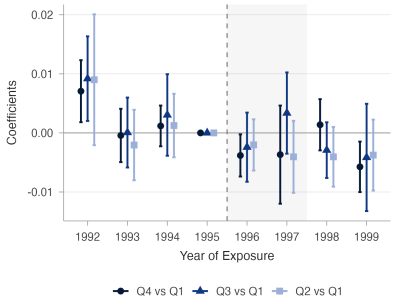
	STEM	Married	Being a mother (female)	Under poverty threshold	Food stamp recipient	Work in the state of birth
	(1)	(2)	(3)	(4)	(5)	(6)
High CNSA × Exposed cohorts	0.002	-0.002	0.010	-0.001	0.000	-0.001
	(0.003)	(0.006)	(0.010)	(0.001)	(0.003)	(0.005)
Observations	1,065,355	1,065,355	514,896	1,007,517	1,065,355	1,065,355
R^2	0.020	0.080	0.063	0.013	0.043	0.028
Dep. var. mean	0.142	0.225	0.260	0.016	0.134	0.693
Est./Dep. var. mean	1.720%	-0.921%	3.718%	-9.236%	0.094%	-0.188%
State-of-birth×Survey-year FEs	✓	✓	✓	✓	✓	✓
Quarter-year-of-birth FEs	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓

Notes: This table presents cohort-DiD estimates with standard errors clustered at the state of birth. High CNSA is an indicator for states in the 4th quartile of baseline CNSA rates. Exposed cohorts are those whose first trimester ended after the March 1996 authorization of folic acid fortification. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. Regressions and dependent-variable means are weighted by IPUMS person weights. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

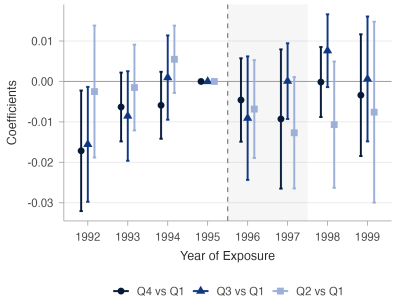
Figure A2: Comparisons across dose groups, event-study plots



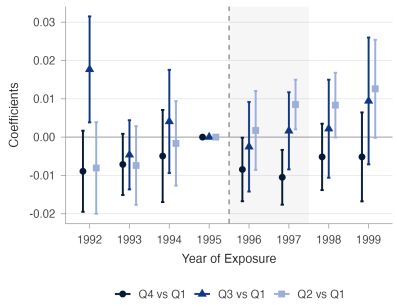
(a) Graduate school enrollment



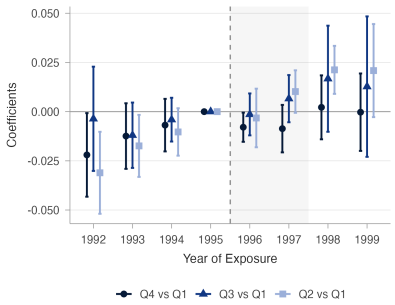
(b) High school diploma/GED



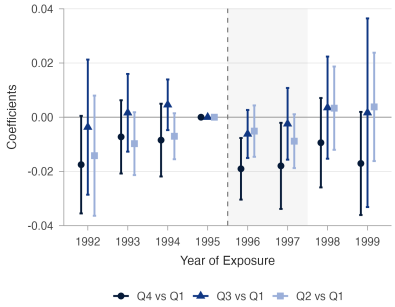
(c) Bachelor's degree



(d) Labor force participation



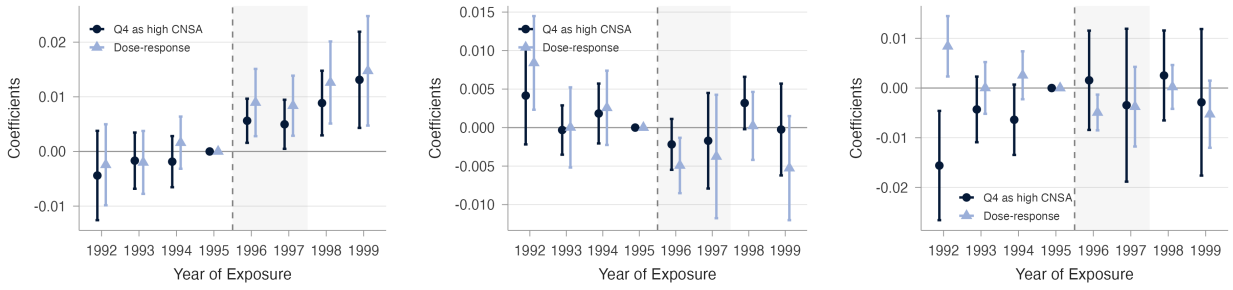
(e) Weekly working hours ≥ 35



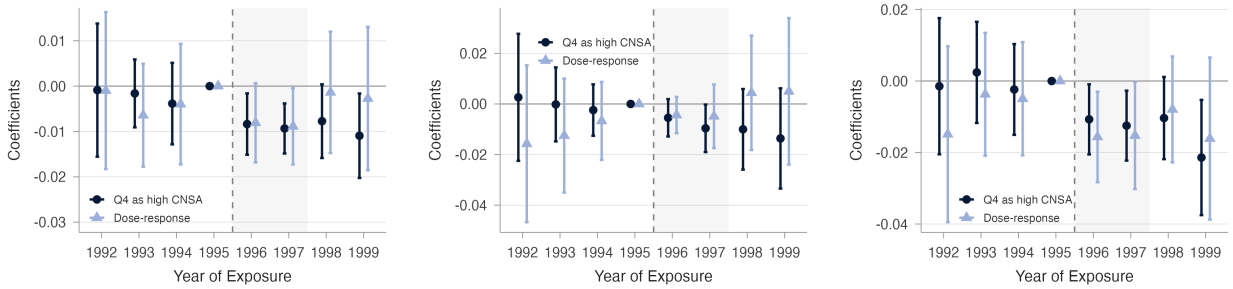
(f) Income (thousands)

Notes: These figures plot event-study estimates with 95% confidence intervals. Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). Regressions are weighted by IPUMS person weights.

Figure A3: Baseline versus dose-response results, event-study plots



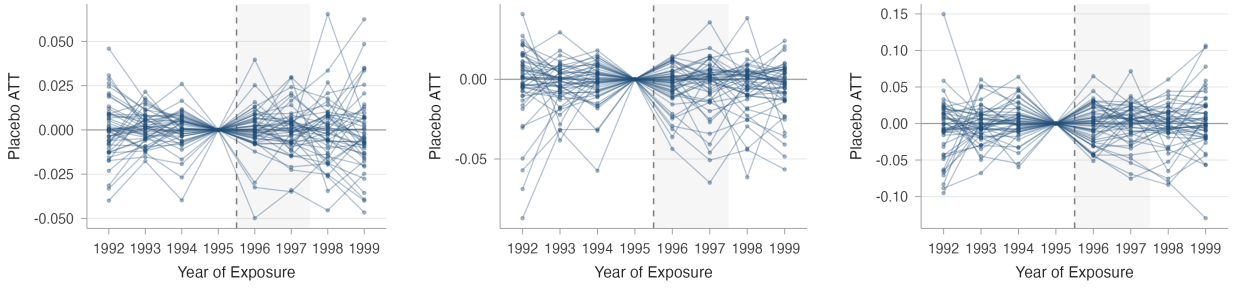
(a) Graduate school enrollment **(b)** High school diploma/GED **(c)** Bachelor's degree



(d) Labor force participation **(e)** Weweekly working hours ≥ 35 **(f)** Income (thousands)

Notes: These figures plot event-study estimates with 95% confidence intervals. Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). Regressions are weighted by IPUMS person weights.

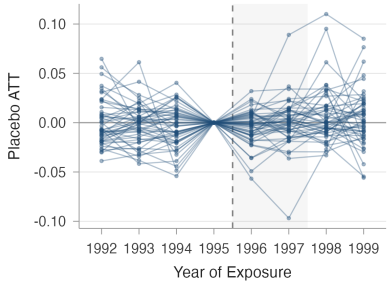
Figure A4: Testing SPT assumption



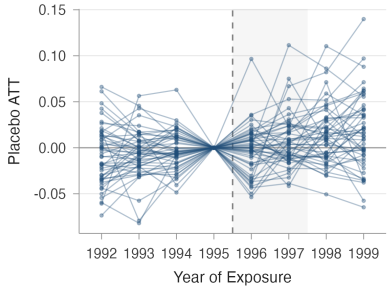
(a) Graduate school enrollment

(b) High school diploma/GED

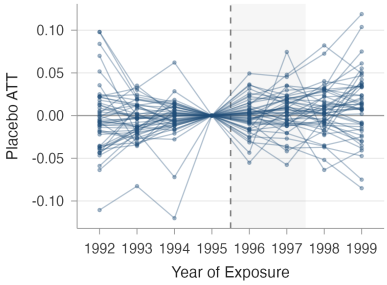
(c) Bachelor's degree



(d) Labor force participation



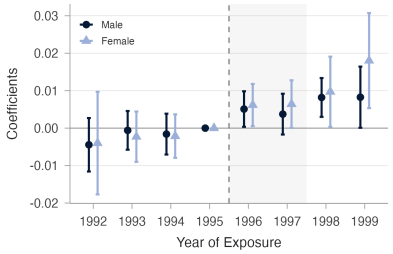
(e) Weekly working hours ≥ 35



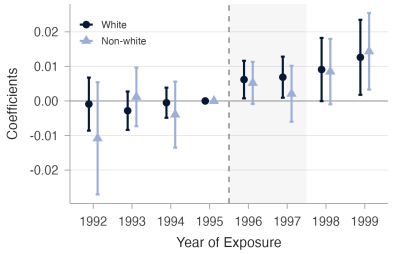
(f) Income (thousands)

Note: In each figure, every line plots the placebo event-study coefficients for one state, where that state's CNSAR exposure is replicated as a "treated" unit while all other states serve as controls. The stacked dataset pairs each replicated (treated) observation with the full original sample (control), and coefficients are estimated via OLS with state×year, quarter-year, and individual demographic fixed effects and HC1 standard errors. The reference year is 1995. The shaded band marks the fortification transition window (1996–1997). Under strong parallel trends, pre-period coefficients (1992–1994) should be near zero across all placebo replications. Controls include Medicaid threshold for pregnant women, mental health expenditure, waiver expenditure, and TANF expenditure.

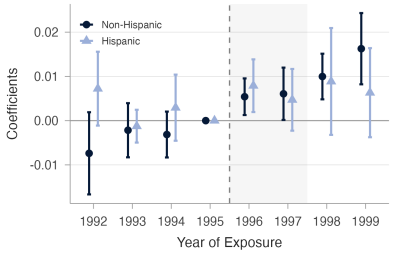
Figure A5: Heterogeneity, educational outcomes, event-study plots



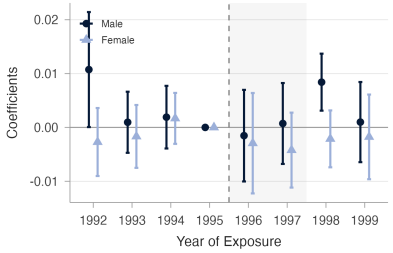
(a) Graduate school enrollment, male versus female



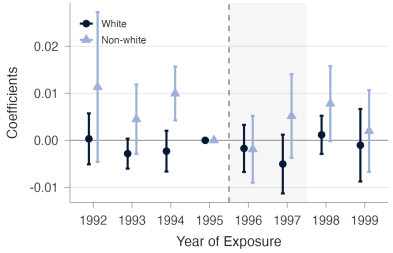
(b) Graduate school enrollment, white versus non-white



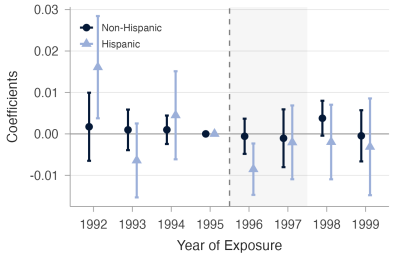
(c) Graduate school enrollment, non-Hispanic versus Hispanic



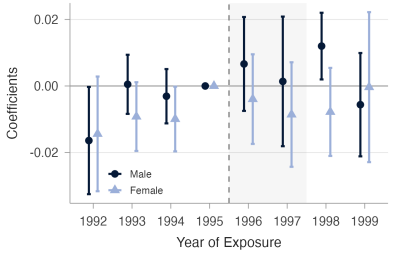
(d) High school diploma/GED, male versus female



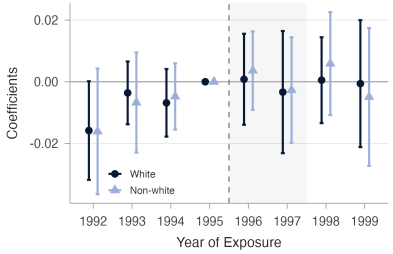
(e) High school diploma/GED, white versus non-white



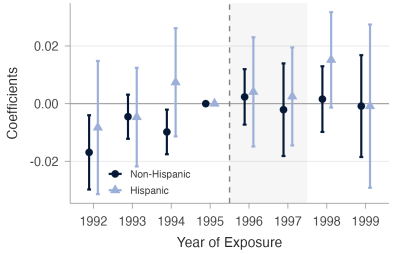
(f) High school diploma/GED, non-Hispanic versus Hispanic



(g) Bachelor's degree, male versus female



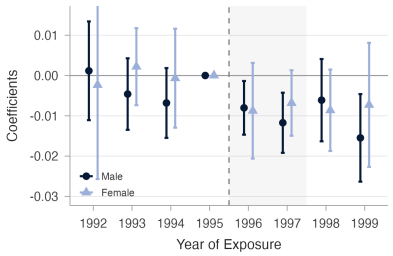
(h) Bachelor's degree, white versus non-white



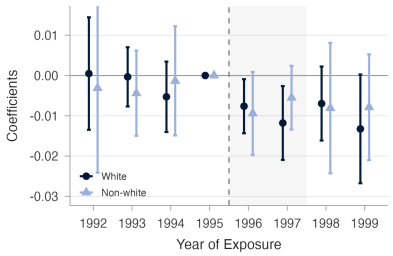
(i) Bachelor's degree, non-Hispanic versus Hispanic

Notes: These figures plot event-study estimates with 95% confidence intervals. Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). Regressions are weighted by IPUMS person weights.

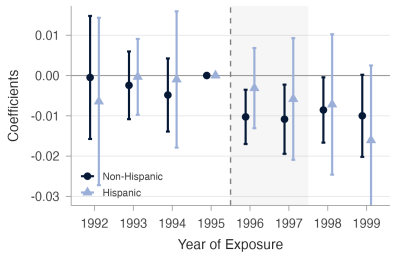
Figure A6: Heterogeneity, labor outcomes, event-study plots



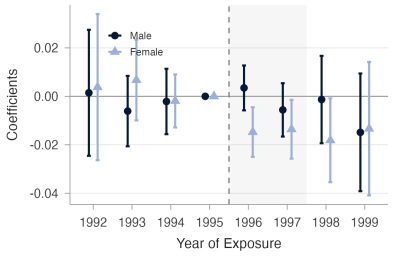
(a) Labor force participation, male versus female



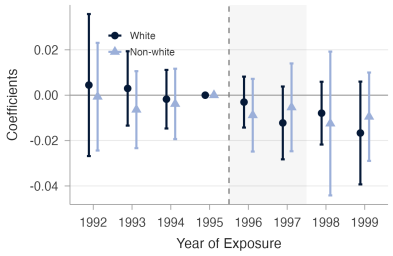
(b) Labor force participation, white versus non-white



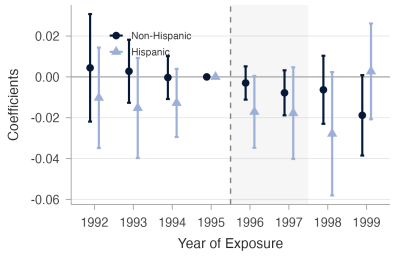
(c) Labor force participation, non-Hispanic versus Hispanic



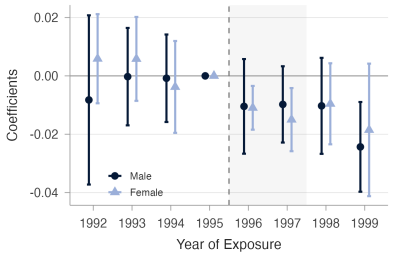
(d) Weekly working hours ≥ 35, male versus female



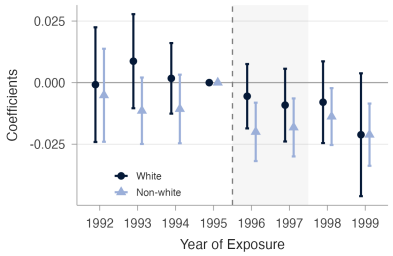
(e) Weekly working hours ≥ 35, white versus non-white



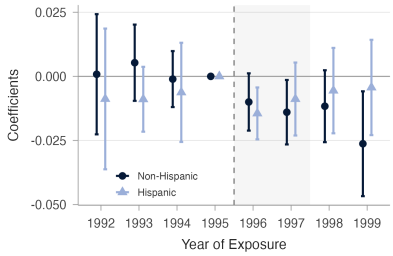
(f) Weekly working hours ≥ 35, non-Hispanic versus Hispanic



(g) Income, male versus female



(h) Income, white versus non-white



(i) Income, non-Hispanic versus Hispanic

Notes: These figures plot event-study estimates with 95% confidence intervals. Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). Regressions are weighted by IPUMS person weights.

Table A3: Effects of folic acid fortification on birth outcomes

	Birth weight (grams)	Low birth weight (1 if birth weight $<$ 2500)	Gestation weeks	Preterm (1 if gestation weeks $<$ 37)
	(1)	(2)	(3)	(4)
High CNSA $\times$ Exposed cohorts	0.074	0.000	0.013	0.001
	(3.013)	(0.001)	(0.018)	(0.001)
Observations	83,631	83,631	83,631	83,631
R^2	0.857	0.793	0.785	0.732
Dep. var. mean	3318.409	0.075	38.840	0.114
Est./Dep. var. mean	0.002%	0.046%	0.033%	1.019%
State-of-birth FEs	✓	✓	✓	✓
Quarter-year-of-birth FEs	✓	✓	✓	✓
Control variables	✓	✓	✓	✓

Notes: This table presents standard DiD estimates with standard errors clustered at the state level. Natality records are aggregated to county-by-quarter-of-birth cells and merged with state-level exposure measures. High CNSA indicates states in the 4th quartile of baseline CNSA rates. Exposed cohorts are those whose first trimester ended after the March 1996 authorization of folic acid fortification. Controls include state-of-birth and quarter-by-year-of-birth fixed effects, along with time-varying covariates for Medicaid eligibility, mental-health parity laws, and welfare reforms. Regressions are weighted by cell birth counts. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

Table A4: Effects of folic acid fortification on maternal characteristics

	Non-white	Hispanic	Age $\leq$ 22 29	Education $<$ college	Unmarried	Indequate prenatal care
	(1)	(2)	(3)	(4)	(5)	(6)
High CNSA $\times$ Exposed cohorts	0.004	-0.002	0.006**	0.003	0.003	0.011
	(0.004)	(0.007)	(0.002)	(0.006)	(0.009)	(0.013)
Observations	83,631	83,631	83,631	83,631	83,631	83,631
R^2	0.985	0.989	0.933	0.949	0.915	0.801
Dep. var. mean	0.210	0.188	0.269	0.549	0.328	0.248
Est./Dep. var. mean	1.989%	-1.259%	2.108%	0.458%	1.035%	4.395%
State-of-birth FEs	✓	✓	✓	✓	✓	✓
Quarter-year-of-birth FEs	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓

Notes: This table presents standard DiD estimates with standard errors clustered at the state level. Natality records are aggregated to county-by-quarter-of-birth cells and merged with state-level exposure measures. High CNSA indicates states in the 4th quartile of baseline CNSA rates. Exposed cohorts are those whose first trimester ended after the March 1996 authorization of folic acid fortification. Controls include state-of-birth and quarter-by-year-of-birth fixed effects, along with time-varying covariates for Medicaid eligibility, mental-health parity laws, and welfare reforms. Regressions are weighted by cell birth counts. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

Table A5: Controlling for place-of-residence-by-survey-survey FEs

	Grad school enrollment		Labor force participation		Income (thousands)	
	(1)	(2)	(3)	(4)	(5)	(6)
High CNSA × Exposed cohorts	0.009***	0.008***	-0.007**	-0.007**	-0.013**	-0.012*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)
Observations	1,065,355	1,065,355	1,065,355	1,065,355	1,065,355	1,065,355
R^2	0.013	0.059	0.013	0.052	0.096	0.163
Dep. var. mean	0.059	0.059	0.843	0.843	0.362	0.362
Est./Dep. var. mean	14.541%	14.327%	-0.865%	-0.862%	-3.679%	-3.281%
State-of-birth×Survey-year FEs	✓	✓	✓	✓	✓	✓
Quarter-year-of-birth FEs	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓
State-of-residence-by-survey-year FEs	✓		✓		✓	
PUMA-of-residence-by-survey-year FEs		✓		✓		✓

Notes: This table presents cohort-DiD estimates with standard errors clustered at the state of birth. High CNSA is an indicator for states in the 4th quartile of baseline CNSA rates. Exposed cohorts are those whose first trimester ended after the March 1996 authorization of folic acid fortification. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms, as well as place-of-residence-by-survey-year fixed effects. Regressions and dependent-variable means are weighted by IPUMS person weights. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

Table A6: Long-run effects on conditional unemployment and occupational quality indices of young adults

	Unemp.	Socio-economic Index, Hauser and Warren	Occ. prestige score, Siegel	Occ. prestige score, Nakao and Treas	Occ. income score	Occ. earnings score, 1950 basis	Occ. education score, 1990 basis
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High CNSA × Exposed cohorts	0.001	12.775	1.180	1.254	0.024	-12.020	-10.456
	(0.002)	(13.284)	(1.626)	(1.647)	(0.062)	(10.082)	(9.948)
Observations	876,876	826,217	826,217	826,217	826,217	826,217	826,217
R^2	0.016	0.055	0.060	0.054	0.035	0.011	0.005
Dep. var. mean	0.062	3750.854	410.225	443.103	27.882	646.357	774.111
Est./Dep. var. mean	0.884%	0.341%	0.288%	0.283%	0.087%	-1.860%	-1.351%
State-of-birth×Survey-year FEs	✓	✓	✓	✓	✓	✓	✓
Quarter-year-of-birth FEs	✓	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓	✓

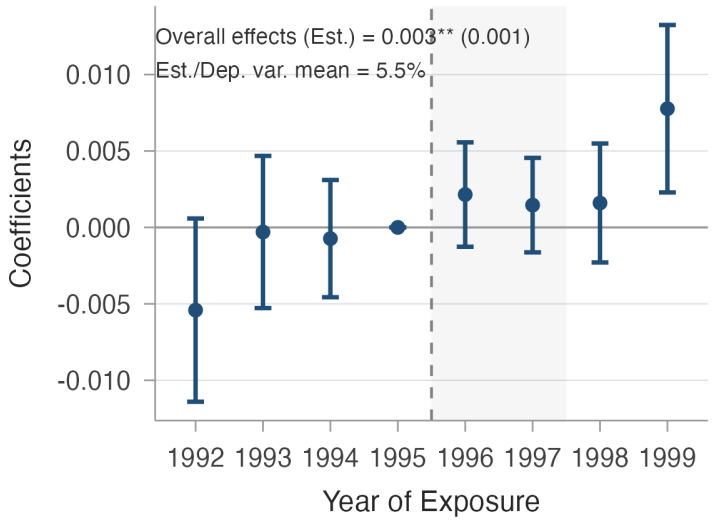
Notes: This table presents cohort-DiD estimates with standard errors clustered at the state of birth. High CNSA is an indicator for states in the 4th quartile of baseline CNSA rates. Exposed cohorts are those whose first trimester ended after the March 1996 authorization of folic acid fortification. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. Regressions and dependent-variable means are weighted by IPUMS person weights. In Column (1), unemployment is conditional on labor force participation. In Columns (2)-(7), occupational quality indices are conditional on being employed. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

Table A7: Long-run effects on conditional unemployment and disability of young adults

	Any difficulty	Cognitive difficulty	Ambulatory difficulty	Independent living difficulty	Self-care difficulty	Vision or hearing difficulty
	(1)	(2)	(3)	(4)	(5)	(6)
High CNSA × Exposed cohorts	0.001	0.003**	0.000	0.001	0.001	-0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
Observations	1,065,355	1,065,355	1,065,355	1,065,355	1,065,355	1,065,355
R^2	0.003	0.002	0.001	0.002	0.001	0.002
Dep. var. mean	0.087	0.059	0.013	0.035	0.009	0.023
Est./Dep. var. mean	1.120%	5.527%	-3.031%	2.041%	6.494%	-6.440%
State-of-birth×Survey-year FEs	✓	✓	✓	✓	✓	✓
Quarter-year-of-birth FEs	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓

Notes: This table presents cohort-DiD estimates with standard errors clustered at the state of birth. High CNSA is an indicator for states in the 4th quartile of baseline CNSA rates. Exposed cohorts are those whose first trimester ended after the March 1996 authorization of folic acid fortification. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. Regressions and dependent-variable means are weighted by IPUMS person weights. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

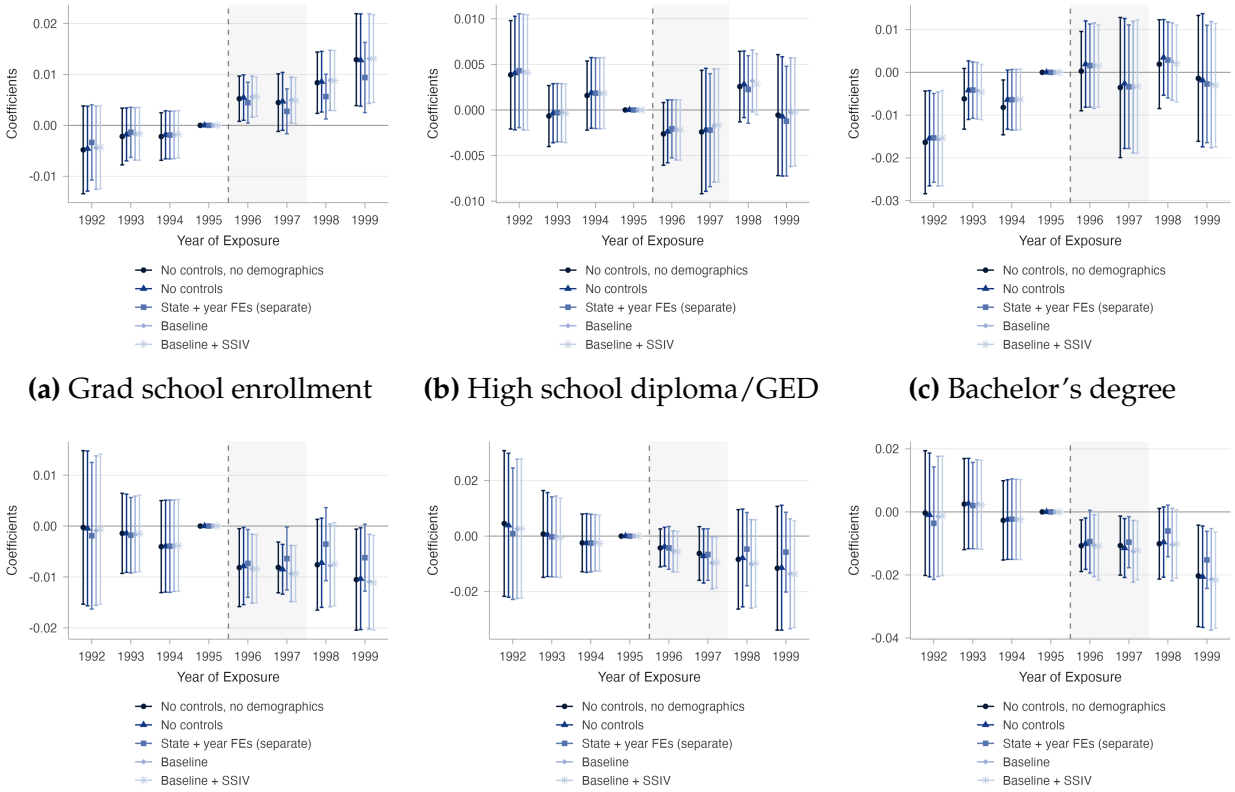
Figure A7: Long-run effects on cognitive difficulty, event-study plots



Notes: This figure plots event-study estimates with 95% confidence intervals. Figure annotations present the overall effect (point estimate, standard error, and the ratio of the point estimate to the dependent variable's mean). Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. Controls include state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects, along with gender, race, ethnicity, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). Regressions are weighted by IPUMS person weights. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.

B Robustness tests

Figure B1: Robustness to alternative model specifications

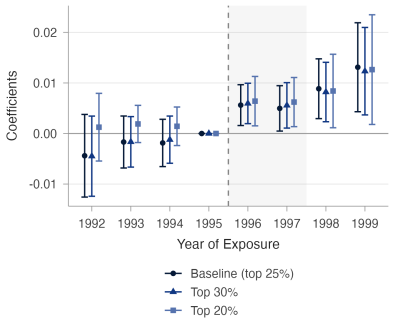


(a) Grad school enrollment **(b)** High school diploma/GED **(c)** Bachelor's degree

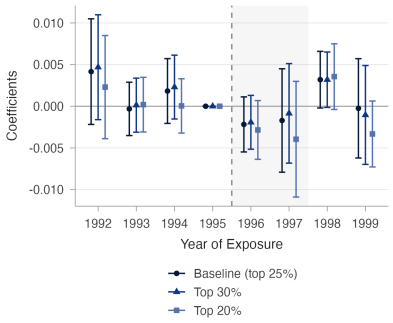
(d) Labor force participation **(e)** Weekly working hours ≥ 35 **(f)** Income (thousands)

Notes: These figures plot event-study estimates with 95% confidence intervals. Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). Regressions are weighted by IPUMS person weights. From left to right: (i) parsimonious model with only state-of-birth-by-survey-year and quarter-by-year-of-birth fixed effects; (ii) preferred specification without confounding policy controls; (iii) preferred specification with fixed effects entered separately; and (iv) preferred specification with an added Bartik-style state-by-year unemployment rate.

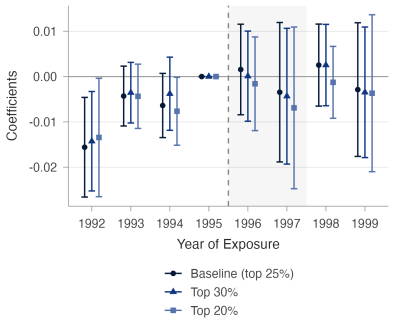
Figure B2: Robustness to alternative exposure measures



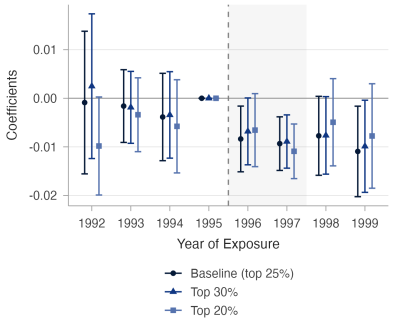
(a) Grad school enrollment



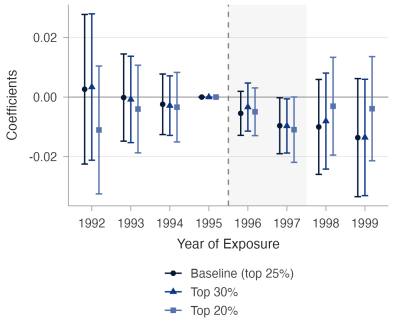
(b) High school diploma/GED



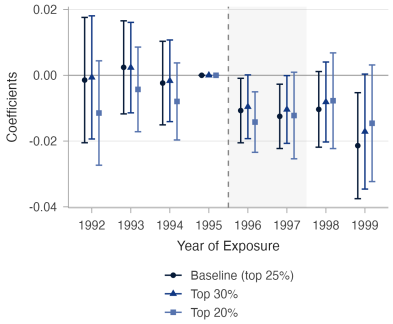
(c) Bachelor's degree



(d) Labor force participation



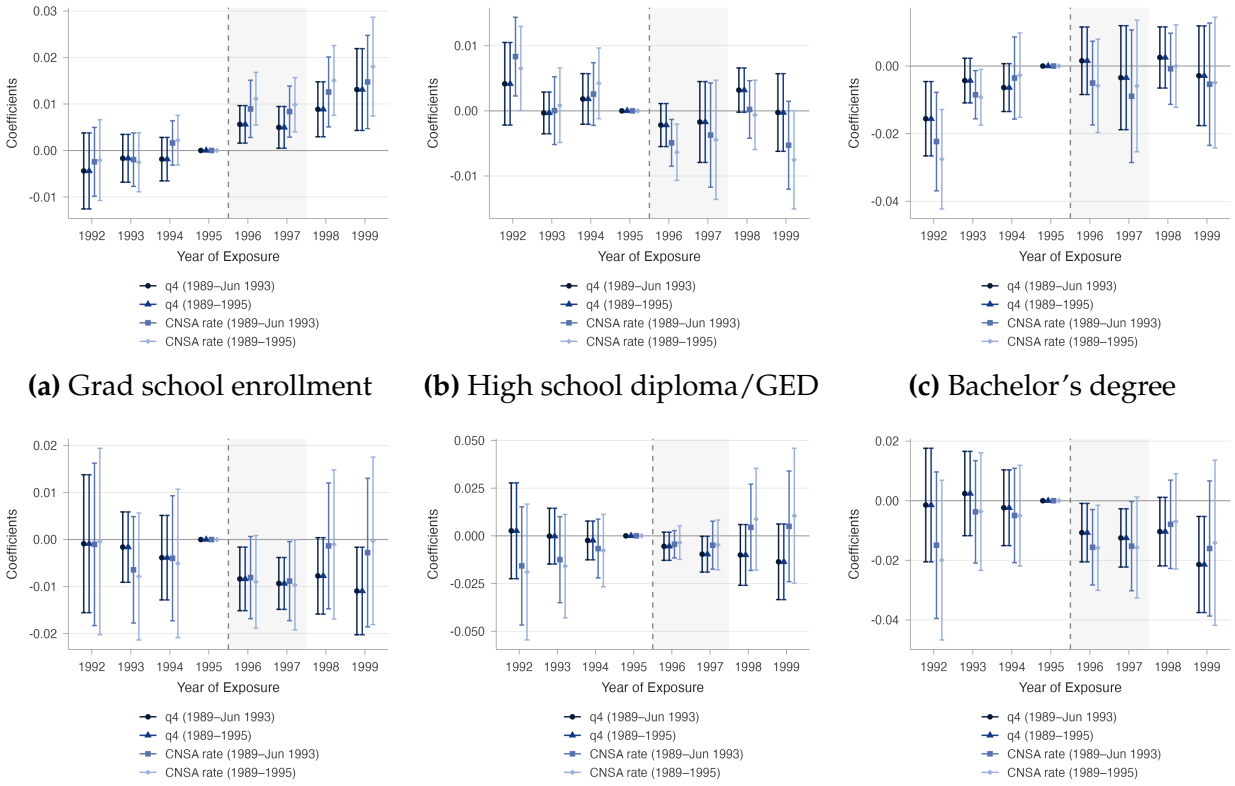
(e) Weekly working hours ≥ 35



(f) Income (thousands)

Notes: These figures plot event-study estimates with 95% confidence intervals. Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). Regressions are weighted by IPUMS person weights. From left to right: (i) top quartile threshold for high CNSA (baseline); (ii) top 30% threshold for high CNSA; and (iii) top 20% threshold for high CNSA.

Figure B3: Robustness to alternative baseline periods

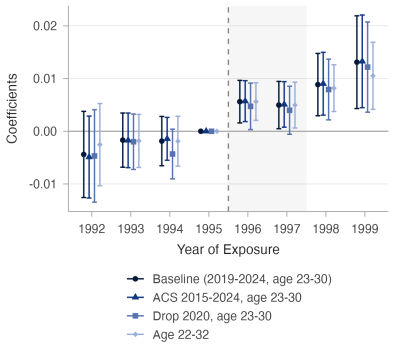


(a) Grad school enrollment **(b)** High school diploma/GED **(c)** Bachelor's degree

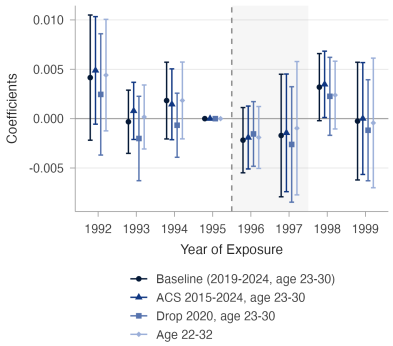
(d) Labor force participation **(e)** Weekly working hours ≥ 35 **(f)** Income (thousands)

Notes: These figures plot event-study estimates with 95% confidence intervals. Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). From left to right: (i) binary exposure (top quartile) with baseline period 1989-1993 June (baseline); (ii) binary exposure (top quartile) with baseline period 1989-1995; (iii) continuous exposure with baseline period 1989-1993 June (baseline); and (iv) continuous exposure with baseline period 1989-1995.

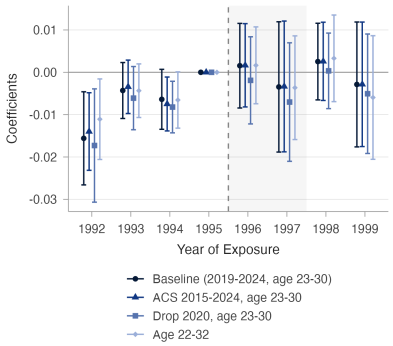
Figure B4: Robustness to alternative samples



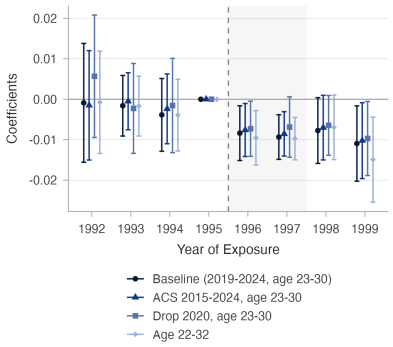
(a) Grad school enrollment



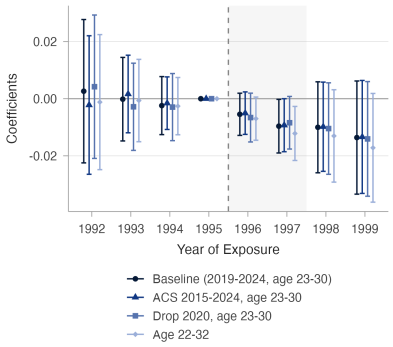
(b) High school diploma/GED



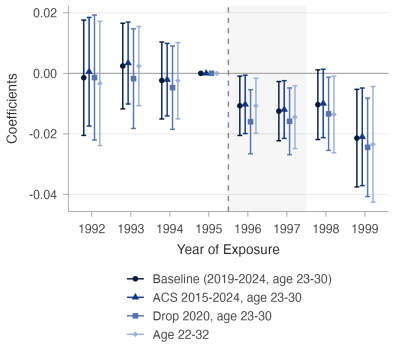
(c) Bachelor's degree



(d) Labor force participation



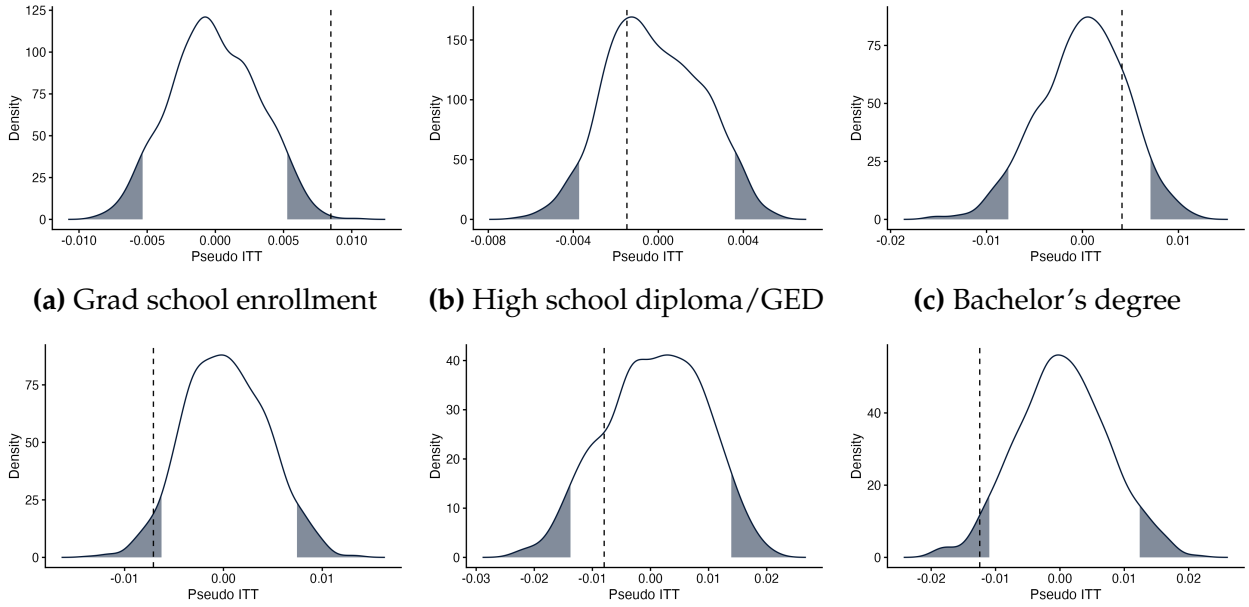
(e) Weekly working hours ≥ 35



(f) Income (thousands)

Notes: These figures plot event-study estimates with 95% confidence intervals. Standard errors are clustered by state of birth. Exposure year groups cohorts into four-quarter bins; for example, 1996 includes cohorts born in 1996Q4–1997Q3. The shaded region denotes partially exposed cohorts (exposed before January 1, 1998). Regressions are weighted by IPUMS person weights. From left to right: (i) PUMS 2015–2024; (ii) PUMS 2019–2024 excluding 2020; and (iii) PUMS 2019–2024 with a wider age range of 22 to 32.

Figure B5: Randomization test

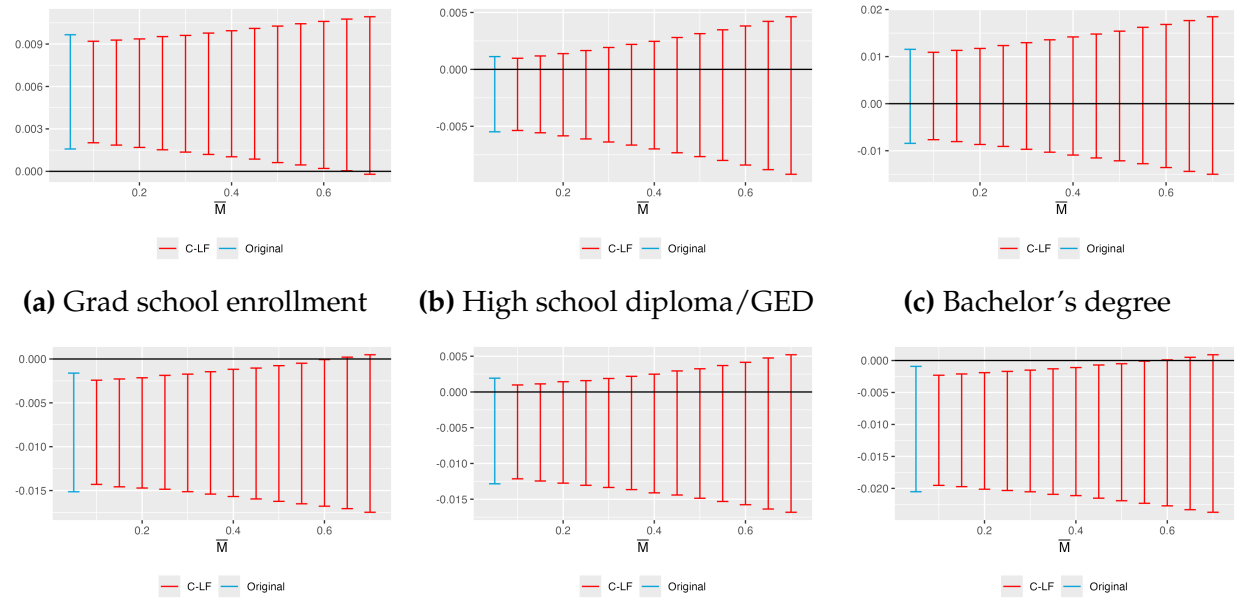


(a) Grad school enrollment **(b)** High school diploma/GED **(c)** Bachelor's degree

(d) Labor force participation **(e)** Weekly working hours ≥ 35 **(f)** Income (thousands)

Notes: These figures compare actual estimates (vertical dashed lines) with pseudo effects from 1,000 random reassignments of fortification exposure, preserving its empirical distribution across states. Regressions are weighted by IPUMS person weights; percentile calculations are weighted by birth counts. Shaded areas represent the 5th and 95th percentiles of the simulated null distribution.

Figure B6: Sensitivity to hypothesized violation of parallel trend assumption

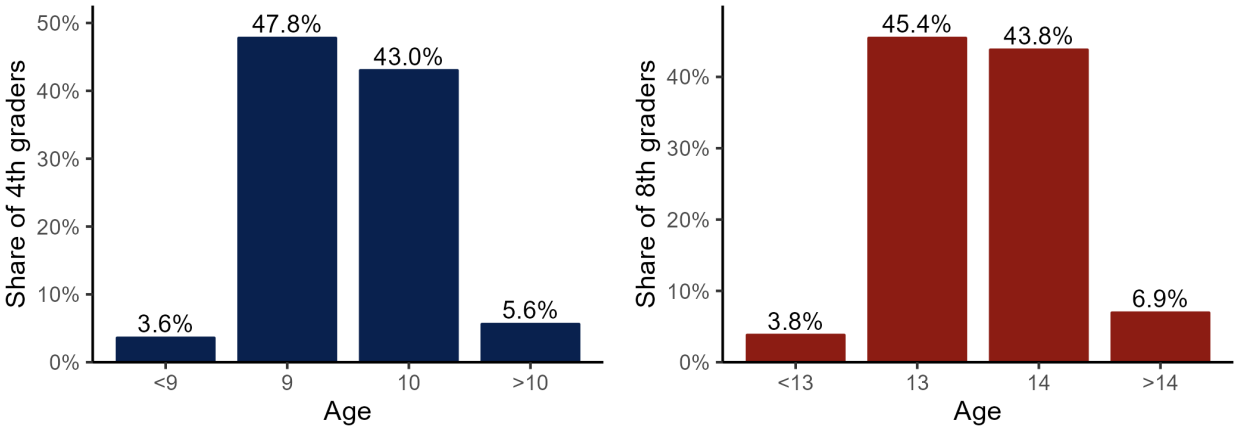


(a) Grad school enrollment **(b)** High school diploma/GED **(c)** Bachelor's degree

(d) Labor force participation **(e)** Weekly working hours $\geq$ 35 **(f)** Income (thousands)

Notes: These figures present Rambachan and Roth (2023) sensitivity tests for the first exposed cohort at the 90% confidence level.

Figure B7: Age distribution of 4th and 8th graders



(a) 4th graders **(b)** 8th graders

Notes: The data is form the ACS PUMS from the same survey year as the public NAEP data.

Table B1: Long-run effects of folic acid fortification on test scores of 4th and 8th graders, smaller assigned year of birth

	Math			Reading		
	Average score	% ≥ basic	% ≥ proficient	Average score	% ≥ basic	% ≥ proficient
	(1)	(2)	(3)	(4)	(5)	
$P(\text{High CNSA}) \times$ Exposed cohorts	4.477***	3.648***	4.599***	-0.811	1.164	-0.987
	(1.083)	(1.058)	(0.857)	(1.078)	(1.477)	(0.616)
Observations	408	408	408	408	408	408
R^2	0.995	0.945	0.939	0.995	0.951	0.945
Dep. var. mean	259.991	75.502	34.361	241.256	69.351	31.305
Est./Dep. var. mean	1.722%	4.832%	13.385%	-0.336%	1.678%	-3.151%
State-of-school-by-year-of-birth FEs	✓	✓	✓	✓	✓	✓
Grade FEs	✓	✓	✓	✓	✓	✓
Control variables	✓	✓	✓	✓	✓	✓

Notes: This table presents cohort-DiD estimates with standard errors clustered at state of school. $P(\text{High CNSA})$ is fortification exposure adjusted for migration. Exposed cohorts are those with likely year of birth after 1996, determined by assigning age 9 to grade 4 and age 13 to grade 8. Controls include state-of-school-by-(likely)-year-of-birth, grade fixed effects, Medicaid eligibility, exposure to mental-health parity laws, and welfare reforms. The unit of observation is state-by-year-by-grade cells. Regressions are weighted by student counts from ACS PUMS. $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$.