Discussion Paper Series

IZA DP No. 18580

April 2026

High Schools and the Uneven Rise in American Opportunity

Alison Doxey
Harris School, University of Chicago

Ezra Karger
Federal Reserve Bank of Chicago

Peter Nencka
Miami University, IZA@LISER and NBER

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18580 April 2026

High Schools and the Uneven Rise in American Opportunity*

Abstract

Between 1850 and 1910, the share of young Americans living in towns with high schools increased from 17% to 46%—the fastest expansion of school access in U.S. history. Using new data on every high school in the United States, we show that this expansion transformed economic opportunities for many young adults but widened class and racial inequalities. We find sharp increases in school attendance rates for high school-aged children in towns that opened a high school relative to children in nearby towns without one. Linking children to adult outcomes, we show that high schools increased women's labor force participation and job quality, while reducing the probability of early marriage and childbearing. Increased access to high school accounts for a third of the increase in women's labor force participation between 1870 and 1930. High schools had the largest effects on children from already-wealthy families, and did not, on average, benefit Black children. While the high school movement substantially narrowed gender gaps in labor market outcomes, it also widened existing race- and class-based disparities.

JEL classification

I26, J24, J16, D63, N31

Keywords

high schools, education, economic history

Corresponding author

Peter Nencka

nenckap@miamioh.edu

* Thanks to seminar participants for their helpful feedback at the Cliometrics Association, Economic History Association, NBER Development of the American Economy Summer Institute, Georgetown University, Mountain West Economic History Conference, Northwestern University, Marquette University, The Ohio State University, Ohio University, Stanford University, University of Chicago, and the University of Wisconsin – Milwaukee. This work was supported by grants from the Institute of Education Sciences and the University of Chicago.

1 Introduction

It is rare for investment in schools to occur rapidly enough for the aggregate consequences to be directly observable. The high school movement of the late 19th and early 20th centuries is an exception. The United States created the world's first widely accessible secondary school system, increasing high school graduation rates from negligible levels in 1870 to 50% by 1940 with large-scale, locally driven investment that saw more than 17,000 high schools built and staffed. At the same time, the country became wealthy and industrialized, shifting away from an agriculture-focused economy as urbanization surged. The composition of the workforce changed: between 1890 and 1940, the percentage of women aged 25 to 44 in the labor force nearly doubled from 18% to 35% (Goldin 2006). Despite massive aggregate gains, Black-white wealth convergence slowed (Derenoncourt et al. 2024) and inequality rose (Chancel and Piketty 2021). Were these aggregate trends driven by the high school movement?

In this paper, we estimate the direct and aggregate effects of the high school movement. This is not a new question: historians and economists have recognized that technological change and new labor-market opportunities during this period coincided with a "race" to educate the population at an unprecedented rate (Goldin and Katz 2008). But isolating the impact of high schools is challenging due to data and econometric limitations. We build a new panel describing every high school in the United States and link both men and women from their childhood hometowns to short- and long-term outcomes in decennial censuses.

To measure the effects of high school access, we use the sharp timing of high school openings and compare cohorts of children born in the same year who lived at different distances from a newly opened school. We combine this distance variation with an event-study framework by tracking each child's age when a high school first opened in their town or city. Our method is analogous to a stacked difference-in-differences design, tracing the relative effects of being near a high school in the years leading up to and following school opening for cohorts who were young enough to take advantage of the new schooling opportunity. We flexibly control for county-by-cohort variation in other amenities, school-opening-year-by-cohort shocks, and time-varying, place-level characteristics. What remains is differential exposure to schools by distance and age at opening, which we show is plausibly orthogonal to other contemporaneous shocks.[1] Our approach builds on an extensive literature that links school construction and school spending to short- and long-run local outcomes (e.g., Duflo 2001, 2004; Aaronson and Mazumder 2011; Navarro-Sola 2021; Andrews 2023; Schmick 2024; Bleemer and Quincy 2025).

We find that the introduction of high schools increased the school attendance rate of affected cohorts and that this increase grew over time, consistent with aggregate attendance trends. We see little increase in attendance among children under age 14, consistent with the idea that high schools did not open at the same time as broader investment in school capacity in affected towns. For 15- to 18-year-olds, we see positive attendance effects that begin in the years immediately after a high school opens. Attendance effects are largest for older students: for example, after two decades, we see an increase of about 28 percent (10 percentage points) in the school attendance rate of 17-18-year-olds.

Linking children to their outcomes as adults, we find that high school entry increased women's (but not men's) formal labor supply at ages 20–28. After high schools opened, both men and women were more likely to enter higher-skilled occupations, such as clerical and professional positions. These shifts translate to increases in predicted occupational income, with much larger effects for women. Consistent with high schools prompting a shift in career choices that required parallel family choices, we observe reductions in the probability that women married and had children in their 20s. When we look ahead another 10 years and examine outcomes for adults at ages 30–38, we see that most labor-market effects persist in sign but attenuate in magnitude, particularly for women. While high schools expanded opportunities for young women, many later left the labor market when they married, consistent with widespread discrimination against women in the

1. In some specifications, we explicitly compare students who were close and farther away from the *same* high school when it opened.

labor market and laws or customs preventing married women from holding certain occupations.

High schools were a transformative institution, and our results explain a large fraction of changes in women's opportunities over this period. For example, aggregate labor force participation (LFP) among women in their 20s increased from 20 to 37 percent between 1870 and 1930. We use a back-of-the-envelope calculation to show that high schools explain roughly a third of this shift. High schools were an important, but not the only, cause of increased female labor force participation during this period. Because the labor force participation rate of men did not change over this period, our results imply that the high school movement played a key role in the convergence of gender roles, particularly for younger adults. Beyond participation, high schools also supercharged emerging occupations, such as clerical work, accounting for almost a third of the growth in female clerical occupations. We use a series of exercises to argue that the direct effect of high schools on attending youth, rather than positive spillovers through changing social norms or negative impacts on crowded-out workers, explains our findings.

The large average effects of high schools mask stark differences by child characteristics. Attendance responses are substantially larger among children from higher-income households, consistent with the large opportunity costs of schooling for lower-income families and potential differences in academic preparation. Changes in adult outcomes are correspondingly concentrated among those with greater parental resources. This is consistent with Parman (2011), who found that communities in Iowa with high schools had lower mobility rates.

While early high schools transformed opportunities for women and reduced gender wage gaps, they *widened* racial wage gaps. High schools were broadly successful at engaging and impacting the white population, but many schools were de facto or de jure segregated, and we estimate precise null effects on attendance for Black children who were near high schools that opened in our national sample. We also find no long run impact on the occupational or demographic outcomes for Black children whose towns built new high schools. While some high schools did enroll Black Americans and may have benefited children who attended, data limitations and limited statistical power may prevent us from observing and measuring this impact. Our aggregate results imply that the early high school movement exacerbated existing racial inequalities in school attendance rates and the labor market.

Our results are robust to a wide range of specification and measurement choices. We show that the distance thresholds used to match high schools to treatment and control towns do not drive our results. Our results are also robust to imposing different rules related to panel balance, constraining comparisons of cities to be within the same high school opening event, controlling flexibly for parental characteristics, limiting our analysis to a subsample of respondents who did not move locations from early childhood to adulthood, and using alternative census linking strategies. We also show that the magnitude of our results is unlikely to be due to chance: we randomly perturb distances to high schools and find consistent null impacts. Lastly, we use an alternative identification strategy. Instead of comparing cities with a high school to places farther away, we use cities that built high schools later as controls for cities that built schools earlier. We find similar results, suggesting that standard selection stories cannot explain the timing, sharpness, and pattern of our results.

Related Literature

This paper makes several contributions. First, we build on the work of economists and historians who describe the high school movement and contemporaneous increases in school spending. Goldin (1998) documents the dramatic expansion of high school attendance and graduation rates between 1910 and 1940, particularly in non-Southern states. Goldin provides one of the first examinations of the correlates of early expansion, showing that investment in high schools was associated at the state level with higher wealth and lower levels of manufacturing. Goldin (1999) extends this work by linking the high school movement to measures of the education wage premium, showing that as high schools expanded across the U.S., the high school wage premium fell by 37 log points from the 1890s to 1939. Goldin and Katz (2008) further explore the interplay between the secondary school movement and the U.S. wage structure. Schmick (2024) and Schmick and Shertzer (2019) focus more broadly on local school spending, examining the determinants and effects of increases in school spending in urban areas during the post-WWI era. We focus on capital investment in high schools and link high school access to both the short- and long-run outcomes of affected children. We also concentrate on women, for whom it has only recently become feasible to study linked child-adult records at scale using the Census Tree Project (Buckles et al. 2025).

Our finding that Black Americans did not, on average, benefit from the early high school boom contributes to work exploring access to and returns to schooling for different subgroups throughout American history. Carruthers and Wanamaker (2017) find that differences in early 20th-century school quality and resources can explain up to 50 percent of male Black-white wage inequality in the South by 1940. Similarly, Aaronson and Mazumder (2011) show that the Rosenwald Rural School construction initiative dramatically reduced racial achievement gaps in the 1910s and 1920s by increasing investment in schools for Black children, and Cascio and Lewis (2024) find that increasing Black teachers' salaries in the South in the 1930s had positive effects on both attendance and educational attainment for affected students. Tyson (2025) finds that in the post-1900 period many high schools available to Black students were focused on technical trades (e.g., agricultural and mechanical training) as opposed to broader skills and that the lack of academic options hindered Black economic progress.[2] We focus on the earliest increase in high school investment, which occurred later in the South and, when it eventually arrived, often excluded Black children.

Our results indicate that women disproportionately benefited from high schools, contributing to the literature on women and schooling during this period. Rury (1991) examines the history of secondary schools after the Civil War and the particular vocational training that they provided to women. Estimating the causal effect of schooling for women has, until recently, been challenging because it is difficult to link women's microdata to their childhood hometowns, as they usually changed their legal names upon marriage. Lleras-Muney (2002) is an early exception, using retro-

2. These studies build on an earlier, large literature that identified gaps in both the quantity and quality of schooling as a potential cause of racial differences in income. (e.g., Welch 1974; Smith 1984; Margo 1986; Smith and Welch 1986, 1989; Margo 1990)

spective data on state of birth to estimate the effects of state-level compulsory school laws (CSLs) on both men and women and find that CSLs increased years of schooling. Li (2025) uses a similar CSL-based approach and modern linking methods and finds that additional schooling increased labor supply and improved "marriage quality" for women. We focus on high school attendance in this paper, which was not mandated by CSLs in most states until later in the 20th century. High school access was also not a state-level source of variation: high school investment varied tremendously at a granular, local level. High schools may also have had a particularly large effect on women's labor-market and demographic outcomes because their mission was partially vocational: developing workforce skills and providing classical college-preparation coursework. Many women attended high school to learn the skills the labor market demanded.

Our work also relates more broadly to modern and historical scholarship that links school construction and school spending to long-term outcomes. Duflo (2001) pioneered this empirical design, using the construction of schools across Indonesia in the 1970s to measure the effects of educational access on wages. A sizable recent literature, summarized in Jackson and Mackevicius (2024), measures the effects of school capital investment on short-run (test score) and long-run (educational attainment) outcomes. Our work provides a historical complement to this literature, focusing on one of the formative expansions of educational opportunity during a period of industrial transformation.

The rest of our paper is organized as follows. We describe the historical background that informs our approach in Section 2. We review our data and sample construction in Section 3. We present and interpret attendance and labor market results in Section 4. We explore additional heterogeneity by race and family income, and discuss mechanisms, in Section 5. In Section 6, we show that our results are robust to a range of measurement and specification choices. In Section 7, we put our results in context with broader trends occurring at this point in U.S. history and interpret the magnitude of our effects. We conclude in Section 8.

2 The high school (building) movement

Our new data and approach build on existing scholarship on the high school movement. This history has, with good reason, focused on understanding attendance and graduation patterns nationwide. But students could only attend schools that existed. We collect new data on the location of each high school built in the United States and show that the high school movement predated the massive enrollment increases in the 1910s that are the focus of much important prior work (e.g., Goldin 1998). The granular nature of our new panel helps us to better map and describe the precise timing of the high school movement, its geographic scope, and its direct impact on children.

The high school movement was informed by past efforts to rapidly expand access to other types of schools in the U.S. The common school movement, led by Horace Mann and other reformers, spanned the 1820s to the 1860s and was successful in spreading state-funded elementary schools in the North. Beyond making school more easily available, the goal of early grade-school reformers was to ensure uniformity of access and quality, as well as rigor (Kaestle 1983). This uniformity and rigor were characterized by the widespread adoption of textbooks such as the *McGuffey Readers* and the introduction of "grades" in larger schools to separate children by age and ability. Reformers highlighted the positive externalities of schooling and argued for local tax support of schooling. This financing model became the basis for the high school movement that followed.

There is some debate over where the first high school in America was founded. If a high school is simply a school that enrolls those aged 14–18, the boys-only Boston Latin School, founded in 1635, has a claim. Its curriculum, however, was much closer to the traditional grammar schools popular in the United Kingdom. These schools and similar private academies focused primarily on college preparation, with little time for applied science or more practical subjects. High school historian William Reese identifies nearby English High School in Boston as the first "modern" public high school — aimed not only at classical Latin and Greek education but at a broader set of topics that appealed to children not preparing for college (Reese 1999). Opened in 1821, English

High School focused on a curriculum of languages, mathematics, and the applied sciences, and was open to all Boston men who wished to attend.

As the common school movement quickly spread the gospel of standardized elementary schools, public high schools also became more common. High school enrollment during the late 1800s did not grow as fast as elementary school enrollment. But, over the 50-year period from 1850 to 1900, we estimate that the number of places (towns or cities) with a high school increased by a staggering 1,200%, from about 730 places in 1850 to nearly 9,700 places in 1900 (Figure A2). This early construction boom predates the better-known increase in secondary attendance and graduation during the 1910s–1930s, and it is central for our empirical approach: we rely on the staggered timing of first high school openings across thousands of places as sharp, place-specific entry events that allow us to compare adjacent towns before and after local high schools were built. High schools continued to expand in the early part of the 20th century — cities opened more schools, existing schools grew, and attendance swelled — but the early wave of high school construction is both important and understudied.

This explosion of school construction sparked fierce debates over school funding. Unlike private schools and academies, which relied on fees, new public schools required local funding and taxes. It was only later that the federal and most state governments got involved. While common school reformers convinced many communities that universal grade school was worth paying for, it was more difficult to convince communities to subsidize education that benefited families who could already afford to have able-bodied teenagers out of the workforce. These financing fights often ended up in court, and a series of legal decisions in the 1860s and 1870s cemented local authorities' ability to levy taxes for secondary schooling. A construction boom quickly followed.

Beyond funding, the next most contentious issue for high schools founded in the late 1800s was what they should teach. Curricular standardization was becoming the norm in lower grades, but in high schools, there was little guidance and few acceptable textbooks. Schools often offered two "tracks" — a classical track with a curriculum designed to prepare students for further study in college, and a "practical" track with applied math, science, and related fields aimed at preparing people to enter emerging "blue-collar" industries. This split was so fundamental that some government reports of the day enumerated the students in each type of high school separately. This bifurcated approach experience was criticized at a high-profile educators' meeting in 1892: the "Committee of Ten" convened by the National Education Association. This committee recommended standardizing the high school curriculum along the lines that we now recognize: a four-year course of study, a focus on both theoretical and applied subjects for *all* students, and yearly courses in English, Civics, and Mathematics. The reports from these meetings were widely read and helped lead to the standardization of high schools, as the common school movement had done for lower grades a few decades earlier.

This standardization set the stage for the rapid growth in attendance observed at the end of the 19th century and especially in the first decades of the 20th century. The high school — newly tasked not just with sending men to college but also with preparing both men and women for the emerging post-agricultural economy — was now an established part of thousands of towns across America. Our goal in this paper is to trace the effects of this early high school movement on the children who experienced it.

3 Data

3.1 High schools

We construct our panel of U.S. high schools from the 1800s to the mid-1900s by combining information from four sources. First, we gather information from censuses of all public and private high schools, collected by the Bureau of Education every 1–2 years from 1873 to 1905 and in 1912. After 1912, the next census of high schools was published in 1951. These censuses contain information on the number of students and teachers, the length of study, and, in some cases, the founding year of each high school.[3] When available, we use founding years to retrospectively construct a consistent panel of high schools across the United States, even for years when high schools are not surveyed.[4]

For later years, we use lists of accredited high schools published by the Bureau of Education every 2–6 years from 1911 through 1944. Accreditation standards varied by state, and public universities used published bulletins to admit local students with diplomas from high schools that met specific criteria. These criteria often included two requirements: (1) that the high school offer four years of study; and (2) that the high school offer at least a minimum number of math, English, and history credits. These lists of accredited high schools contain no information on founding dates or attendance, but they confirm the existence of public and private secondary schools in particular cities and towns each year.

To complement the founding dates from the earlier high school censuses, we use *Patterson's American Educational Directories*. These provide names of high schools in each county for seven years between 1906 and 1924. The Patterson's directories contain establishment dates for reporting high schools. However, these directories were less comprehensive than the contemporaneous censuses of high schools produced by the Bureau of Education.[5]

To supplement our data on private high schools, we collect information from the more recent Private School Universe Surveys produced by the National Center for Education Statistics. These surveys include the founding dates of reporting schools. We augment our historical panel with the founding dates of high schools reported in the 1989 and 1995 academic year surveys to ensure coverage of long-lived private high schools that may have been missed in earlier sources.

We hand-digitize each source except the more recent Private School Survey, which is already

3. The founding year was only collected in a handful of the later censuses, including in 1903.

4. A limitation of this reliance on retrospective data is that we do not observe high schools that opened and closed between school censuses. In our empirical analysis, we focus on whether a high school is present in a town or city. A town or city rarely loses access to a high school once built, so we do not believe this is an important limitation for our analytical purposes.

5. We use Patterson's directories for the 1906, 1908, 1912, 1913, 1914, 1920, and 1924 academic years, bridging the gap between available Bureau of Education censuses of high schools and the later comprehensive lists of accredited high schools.

available in machine-readable form. We combine these four sources into a single panel of high schools, yielding 360,000 school-by-year observations. The median observation in our panel was published in 1922, but the critical data for our analysis come from the 1889–1905 high school censuses, which are available annually.

To prepare this panel for analysis, we first identify transcription errors in the names of towns and cities and correct them by looking for places that appear uniquely in the panel, differ in spelling by one letter from other place names in the panel, and do not conflict with available years. To geocode each place, we use the Google Maps API to link each high school's town or city name to a latitude-longitude representing its location. This API is our preferred geolocation tool because it can often correctly geocode locations with spelling errors. We then collapse the data to the place-year level, calculating attendance for local schools and the earliest reported founding year for public and private schools.

At the end of this process, we have a panel of 23,694 places with information about the longitude and latitude of the place where we observe that high school. Figure 1 maps the spread of high schools over time. Towns and cities with high schools by the 1800s are concentrated in New England and the Midwest. High schools expanded later in the South, and towns and cities in the West also lagged New England in schooling investment. Figure A1 shows that despite these general patterns documented in past work, measures of aggregate state-level high school access mask substantial heterogeneity within states. For example, comparing Panels A and D, we see that the average town in Massachusetts had an operating high school much earlier than the average town in Tennessee. But *within* each state, there is substantial variation in the timing of high school access. This within-state and within-county variation is particularly evident in the Midwest, as shown in Panels B and C for Ohio and Wisconsin.

In Panel A of Figure A2, we plot the increase in the number of towns and cities in the United States with at least one high school between 1800 and 1950. We show the cumulative number of cities with a public high school and 'any' high school, which includes private high schools.[6] While some towns and cities built high schools in the early to mid-1800s, we see the largest increase in high schools between 1880 and 1910, when the number of towns with high schools more than tripled.

A key feature of our data is that its quality is highest precisely when the "first wave" of high school entry occurred. The Bureau of Education's school-level censuses of high schools are the best available source for identifying founding years and (in many years) enrollment, and are available at high frequency through the early 1900s. After 1910, we no longer have access to higher-quality high school census data. The flat growth rate in the number of towns with a high school between 1910 and 1915 is driven by data limitations. The subsequent upswing in the number of towns with a high school in our panel is driven by the large number of accredited high schools that, for the first time, reported their existence to the Bureau of Education in microdata in the 1920s and 1930s. Because of these changes in data quality, in our main results, we focus on variation in high school construction before 1909 and exclude cities that established high schools after 1909.[7] To some extent, this data restriction implies that we are looking at the subset of "early-moving" towns that built high schools; the median year of first high school among cities and towns is coincidentally 1909 in our full panel. That said, we do capture much of the explosive growth in high school attendance that occurred from 1910 to 1930 in our analysis. This is because we allow the effects of "first high school" access to vary over time, and as we show in Section 4, place-level attendance spikes after a high school is built but continues to grow in the decades after entry.[8]

6. Note that we plot the earliest year when a high school could have existed in a town, according to our panel. So, for example, because we have lower-quality data between 1905 and 1910, we assume that each town that reported a high school in 1910 had access to that high school as of 1906—the first year in which our panel indicates a high school could have existed in those towns. This approach creates discontinuities in the growth rate in Figure A2.

7. Our main analysis sample marks locations as 'ever-treated' if they contained any high school before 1909 and 'never-treated' if they did not establish a high school before 1950. We include cities that established a high school before 1850 in our analysis sample as 'always-treated' locations, and they can contribute to the estimation of longer-run post-treatment coefficients. However, since 1850 is our first census year, they do not contribute to pre-treatment estimates. In Section 6, we show that our results are similar if we only include more balanced observations that have both pre- and post-periods within the 1850–1940 timespan.

8. Cities and towns that built high schools later expanded them or added additional schools. In addition, schools

In Panel B of Figure A2, we plot the total enrollment of students in high schools in the years of our data with valid enrollment information. Consistent with prior work by Goldin (1998), who reports Bureau of Education aggregate graduation and enrollment information from 1890 to 1970, we find a relatively linear increase in enrollment rates from 1890 to 1910.[9] In years when we have enrollment information aggregated from the high school panel and Goldin (1998) reports aggregate enrollment rates, our reported enrollment rates are similar. For example, in 1910, Goldin reports a high school enrollment rate for 14–17-year-olds of around 15%, and there were roughly 7 million 14–17-year-olds in the U.S. at that time, implying an enrollment of 1.05 million students. Aggregating our high school panel, we estimate that 1.06 million students attended high schools in 1910.[10]

3.2 Census data

We use the complete count historical decennial U.S. censuses to track attendance and adult outcomes (Ruggles et al., 2024). We use restricted census data for this project, which includes respondents' names. However, as we discuss below, most of our analysis does not rely on knowing respondents' names, and our main results can be fully replicated using public full-count census data from IPUMS.

Attendance and household characteristics

We measure a first-stage effect of high schools on school attendance using contemporaneously reported attendance data from the 1850 to 1930 decennial censuses. Each census after 1850 asked whether children in a household attended school in the previous year, though the format and reporting patterns of those questions changed over time. For example, the 1910 Census Enumeration Instructions gave the following directions:

---

were not always at full capacity from the year that they opened.

9. Goldin (1998) finds a roughly linear increase in graduation rates from 1890 through 1900 and a linear increase in attendance rates from 1900 through 1910 when the Bureau of Education begins to report this information.

10. Note that in Figure A2, we report this total in 1912, which was the year of the Bureau of Education Report, but which refers to the 1910 school year.

> Write "Yes" for any person who attended school, college, or any educational institution at any time since September 1, 1909, and "No" for any person of school age—5 to 21 years—who has not attended school since that date. For persons below or above school age, leave the column blank; unless they actually attended school. (United States Bureau of the Census 1910)

This historical instruction to census enumerators creates difficulties for researchers because it does not define what constitutes "attending" school or what constitutes an "educational institution." In some years, enumerators gave explicit instructions that a respondent who attended only Sunday or evening school should not count as attending school.[11] However, it is unlikely that this instruction was always followed in the years when it was given, and in many years, only the general instruction was given. Much of the attendance that we observe in the censuses, then, is likely not full-time attendance in traditional schools. And even for 15-16 year olds who report attendance and actually attend full-time school, reported attendance might reflect time spent in common schools (shared with other students) rather than in traditional high schools.[12]

For these reasons, we cannot use the census to accurately count the number of children enrolled in high school.[13] However, for our purposes, the attendance question can still be useful. We will look for *jumps* in the probability of reporting attendance when high schools open nearby. So long as the error induced by census attendance measurement is not correlated with the timing of high school entry or the distance to a new high school, our estimates will be an informative measure of how high schools changed attendance decisions.

We also show attendance results separately for children who do and do not report an occupation in the census. Following Goldin and Katz (2011), we define "full-time" attendance as reporting

11. This is the case in the 1870 enumeration instructions, which explicitly noted that the field was "not intended to include those whose education has been limited to Sunday or evening schools." (United States Department of the Interior 1870)

12. Both of these points are made in Goldin and Katz (1998) in the context of the 1910 and 1920 population censuses. For this reason, they note that those censuses "probably overstate the proportion of youths in secondary schools".

13. In much prior work on the high school movement, the focus has been exactly this: getting accurate attendance and graduation numbers. The census limitations outlined here suggest that one cannot do that from the censuses alone. It is preferable to use administrative state and federal data from schools themselves, as in much of Goldin and Katz's work.

schooling but not an occupation in the census and "part-time" as listing both attendance and an occupation. As Goldin and Katz (2011) note, this is an imperfect proxy for attendance intensity. Young people might report working over the summer or holding a part-time job while still attending school full-time. But it is a useful proxy. *A priori*, high schools could increase both types of attendance and also shift students from less formal types of schooling that are likely to be 'part-time' (e.g., Sunday school) to 'full-time' status.

To assign children to high schools, we combine the high school and census data. Much of the prior work focuses on state-level measures of high school investment, or in some cases, county-level measures. We are interested in local high school access, so we focus on links at the *place*-level. We make geographic links using the Census Place Project (CPP), which converts census location strings into consistently geocoded cities, towns, and unincorporated places (Berkes, Karger, and Nencka 2023).[14] We focus on small and mid-sized cities and rural areas. For these locations, a nearby high school is a plausibly salient treatment. We exclude larger cities (e.g., New York City) where within-city distance to a high school is a strong determinant of practical high school access.[15]

In addition to attendance and geographic location, we observe the household composition of each child in our sample. We record information on their parents' labor force participation status and imputed occupational income. We do this separately for mothers and fathers for use in heterogeneity and robustness exercises.

Adult outcomes

We link individuals we observe in childhood to their adult outcomes in subsequent years. Unlike in some modern administrative datasets, linking is a nontrivial task because there are no unique identifiers across census years. Recent innovations in the historical record linking literature allow us to track both men and women over time and across places.

14. Throughout the rest of the paper, we will refer to these places as "cities," "towns," and sometimes simply as "places." For our purposes, all of these terms refer to this same type of sub-county location described above, which the Census Place Project assigns to an exact latitude and longitude.

15. We cannot consistently geocode the addresses of either high schools or census respondents.

Our baseline analysis uses links from The Census Tree Project (Buckles et al. 2025). The Census Tree provides decade-to-decade links generated through several methods. The initial set of links comes from user-provided records on FamilySearch, a genealogical website. Buckles et al. (2025) use these links as training data to generate additional links with a supervised machine learning algorithm. They supplement these links with connections from the Census Linking Project (CLP; Abramitzky et al. 2022), the IPUMS Multigenerational Longitudinal Panel (MLP; Helgertz et al. 2023), and “hints” generated by algorithms on the FamilySearch website. Candidate links from these methods are adjudicated and provided as decade-to-decade crosswalks.

Compared with other potential linking methods, Census Tree links are notable for their strong coverage of women. Deterministic, name-based approaches like the CLP work well for men and can work well for linking older women. However, since women often change their names when they marry, these methods cannot be used to link girls in childhood census records to their adult outcomes. The IPUMS Multigenerational Longitudinal Panel includes high-quality links for a subset of women who stay in the same household as a linked male relative (e.g., a father or a husband). Similar to CLP-style methods, these links produce a selected sample when used to link women to their childhood homes. The genealogical data in the Family Tree dataset provides a unique and powerful way to link large numbers of teenage women forward, even as they change names, move households, and start careers.[16]

3.3 Sample construction

To construct our analysis sample, we begin with the universe of children aged 10–18 in each decennial census from 1850 to 1930. We link each of these children to the CPP to capture locations and record information on their attendance and parents, as discussed above. We then create a sample of adults aged 18 or older for each census from 1860 through 1940. We capture all available labor force and demographic data for these adults. Returning to the childhood records, for each

16. A limitation of many linking methods, including the Census Tree, is that they include fewer Black individuals.

census decade, we attempt to link each child forward to their age-20 and age-30 outcomes using the Census Tree links. For example, a child who was 14 in 1860 could be found when they were 24 in 1870 and 34 in 1880.[17] We also link children aged 10–18 in 1860–1930 backward one decade to their younger childhood observations to observe their location and characteristics when they were ages 0–8; we use these links in robustness exercises.[18]

Having constructed a sample of all 10–18-year-olds linked to their adult outcomes in their 20s and 30s, we match each child to their nearest high school that ever opened, often in their exact town or city of residence. We take all children and find the nearest place within 0.5 miles that constructed a high school and we record the date that place received its first school. In the rare event that there are multiple places within that half mile, we take the place that built a school first. This distance captures fuzziness in geocoding methods across sources. If the closest school is more than 0.5 miles away, we record the opening date of that high school. Our baseline sample uses children within 0.5 miles of a town with a high school as the treatment group and children more than 3 miles away from a town with a high school as the control group.

For the reasons discussed above, we limit our analysis sample to places whose nearest high school opened before 1909. We also drop the largest twenty-five cities in 1860—these are places where city-level access to a high school is likely not the relevant margin of analysis. Last, we exclude children who live in places that we do not observe in at least six of the eight possible decennial censuses between 1850 and 1930 (recalling that the 1890 census microdata does not exist). We do this to avoid including the large number of towns founded only in the early 1900s in our sample; they often built new high schools soon after being established, but we have no way to verify what life was like in these towns before the high school arrived.[19]

Table A1 shows summary statistics for the sample in full and split by gender. On average,

17. Since the last full-count census currently available in a high-quality machine-readable format is the 1940 census, we do not link forward to ages 40 and 50 outcomes because of sample loss.

18. We use a base sample of 10–18-year-olds to maximize our sample at critical high schooling years.

19. While we exclude these cities from the census outcome analysis, we do not exclude the *high schools* themselves from our data. Those high schools can be linked to other nearby places.

people lived close to a high school: the average distance between the city where children lived and the nearest high school is only 1.41 miles. This motivates our empirical strategy of comparing people who live directly in a town with a high school to those who live nearby in a town without a school. Reported enrollment is high; over 30 percent of students aged 17–18 report attendance. As discussed above, this is an overestimate of the actual high school attendance rate. Men are more likely to be in the labor force and in professional occupations, and women are more likely to be in clerical occupations. Women's labor force participation falls from their 20s to their 30s, a pattern that will also appear when we estimate the impact of high schools on women's occupational choices.

4 Attendance and labor market results

This section describes our main results, proceeding in two steps. First, we estimate the effects of high school entry on contemporaneous school attendance. This analysis is a "first stage" and tests whether local high school establishment affects attendance. Second, we use our linked sample to estimate analogous models for the labor-market and demographic outcomes of older adults, whose childhood locations and access to high school we identify from the census links described in Section 3.

4.1 Methodology and treatment definition

To estimate the effects of high school entry on attendance, we specify an event-study model analogous to a stacked difference-in-differences approach. We combine information on the precise timing of high school entry and students' proximity to a given high school when it opened. Students who live close to a high school when it opens are more likely to attend, and they can't attend a high school before it opens. Combining these two sources of variation, we ask: For students from the *same birth cohort*, whose nearest high school opened in the *same year*, how do outcomes differ between those living with a high school in their town vs. their peers living in the same county but in a town a few miles away from the high school that opened? Table A2 shows the distribution of decade openings for high schools and the average distance to a town with a high school for children in our treatment and control samples. By construction, the average distance between the town of residence and the nearest high school town for our treatment group is nearly zero. The control group is, on average, five miles away from their nearest high school. In robustness exercises, we vary both the treatment and control group distance thresholds to help trace out a more flexible picture of how high schools impact attendance and long-run outcomes. Turning to the high school opening by decade panel of Table A2, most of the variation that we use comes in the mid to late 1800s.

Table A3 shows average adult 1860 labor market characteristics for our treatment and control cities, split by treatment timing groups (pre-1860, 1860-1890, and 1890-1909). For the pre-1860 high school columns, these statistics are measured after high schools were built. For the other cohorts, their columns show pre-high school differences in 1860 covariates. Places that received high schools were larger and had higher-earning workforces than places without high schools. Given the historical background and prior work on the high school movement, this is unsurprising: high schools were funded by taxes and were likely most attractive to communities that saw their potential.

Our identification strategy does not require that high school locations be completely unrelated to children's potential outcomes. This is implausible. Instead, we use the sharp *timing* of high school construction interacted with variation in which cohorts were the appropriate age to attend high school to argue that, in a narrow window, the high school opening is the only intervention that sharply changed outcomes exactly for cohorts who were now able to attend. A potential confounder would have had to affect children who could attend high school but not their slightly older peers in the same place, and it would have to have had this effect on average across thousands of towns that built schools in different decades.

To fix ideas, Figure A3 shows a simple version of our identification strategy. We take a set of 18-year-olds in each decennial census and calculate their school attendance in year t and their occupational outcomes in year $t+10$.[20] We then take the unconditional mean of these outcomes separately for our treatment and control cities in the years leading up to and after the nearest high school opening. For control cities, this is a high school that opens at least 3 miles away; for treatment cities, within 0.5 miles. For all outcomes, we see relatively flat unconditional pre-trends in the years leading up to high school entry. After high schools open, gaps emerge: high school enrollment increases more in treatment cities, and the probability of working in professional and clerical occupations increases.[21]

To formalize this approach and use all birth cohorts in our data[22], we estimate an analogous regression specification that uses age at high school entry as our treatment timing variable. Our estimating equation is:

$$Attend_{ictk} = \sum_{a=-30}^{60} \beta_a (Treated_i \times \mathbf{1}[\text{age} = a]_{ictk}) + \lambda_b + \delta_c + \eta_t + \gamma_{ob} + \phi_{kb} + \varepsilon_{ictk} \quad (1)$$

where $Attend_{ictk}$ is an indicator for person i, living in childhood city c, in year t whose nearby high school opened in year k. $Treated_i$ is an indicator equal to 1 if c is ≤ 0.5 miles from a place with a high school and 0 for children more than 3 miles away from a high school. The indicator variable $\mathbf{1}[\text{age} = a]_{ictk}$ tracks person i being age a when a school opened in year k. Next, λ_b, δ_c, η_t, and γ_{ob} are birth-year, city, year, and county-by-birth-year fixed effects, respectively. We also control for school opening year $\times$ birth-year fixed effects (ϕ_{kb}). These fixed effects adjust for trends common to a birth cohort exposed to a school that opened in a specific year. The coefficients of interest

20. These occupational outcomes are calculated by linking the children to wherever they live at time $t+10$.

21. Interestingly, high school "attendance" is higher in the control cities. As discussed in Section 3, the high school attendance measure captures significantly more enrollment than administrative high school data predict, suggesting it captures part-time and informal schooling options, such as Sunday schools. This appears to be more common in our control cities. We return to this and separately estimate the effect of full and part-time enrollment in the next subsection.

22. Censuses are taken every 10 years, so we cannot observe high school enrollment for all children. However, we can estimate for each child their age at high school entry and examine long run-outcomes.

are the vector β_a, which track attendance effects as a function of the age when each high school opened in a town relative to the same trends in places farther away. We cluster standard errors by the nearest high school, since each high school can be matched to multiple cities.[23]

Each coefficient β_a captures a treated–control difference for cohorts defined by their age when the nearest high school first opened. Intuitively, we are comparing children from the same birth cohort, matched to high schools that opened in the same year, and living in the same broader area, but who differ in whether they lived in the high-school town (treated) or a nearby town several miles away (control). In the event-study figures, the "partially treated" region corresponds to cohorts who were already in their mid-to-late teens when the school opened (and therefore could only attend for part of the traditional high school window), while the "fully treated" region corresponds to cohorts who were young enough at entry to potentially complete the full course of secondary study.[24] Under the identifying assumption that no other shocks change discontinuously at the same event time in a way that differentially affects the treated group relative to the control group, post-opening movements in β_a trace the causal effect of local high school access.

The intuition of Equation 1 is similar to a standard stacked event study. To illustrate, consider attendance outcomes for children aged 17–18. If children in the census turned 25 when a high school first opened in their childhood town, it would be impossible for that high school to affect their attendance when they were 17–18. The same is true for children aged 35 when a high school opened in their childhood town. If we see an increase in the predicted probability of attendance when a high school opened at age 25 compared to 35, that is a pre-trend: it suggests that some other characteristic of the town was changing across those cohorts, affecting children's attendance rates.

23. Results are similar when we cluster at the childhood place level.

24. Defining which ages should be affected by a high school is not as straightforward as it is today. If a high school opens during later teenage years, children could be partly treated by the school. If the school opens before they are 14, children would be fully treated because they can attend the high school for the full range of traditional schooling years. Some children older than 18 might attend high school, especially when high schools were first built and had fewer age restrictions. We estimate Equation (1) using three-year age bins with an omitted category of young adults who were 21–23 when schools opened.

Our distance-based estimator differs from a standard two-way fixed effects estimator both in how we define treatment and control groups and in our estimation strategy. High schools expanded rapidly, and most Americans had access to one in their county by the early 20th century. But, as we will show, the exact *distance* to a high school matters for attendance, and idiosyncratic differences in these distances and the timing of high school entry provide sharp variation to estimate the impact of high schools. Our strategy avoids concerns that standard two-way fixed effects models can be biased by "forbidden comparisons" between early- and later-treated units (Goodman-Bacon 2021; De Chaisemartin and D'Haultfœuille 2020). Unlike in a standard event study, event time is well-defined for both the treatment and control groups, since we match each to its nearest high school. Rather than comparing early- and late-treated groups, we rely on clean treatment-control comparisons within the same birth cohort and school-opening windows. In robustness exercises (Section 6), we will also estimate models that constrain comparisons to be within the *same* high school opening event by including high school-by-birth-year fixed effects.[25]

4.2 Attendance results

Figure 2 shows the results of estimating Equation 1 on school attendance. We show results separately for children aged 11–12, 13–14, 15–16, and 17–18. We normalize pre-period coefficients to the average of pre-treatment estimates. For all ages, we see flat pre-trends before high school entry; towns that did and did not establish high schools were not on differential enrollment trends before entry. After high schools are founded, we see no increase in attendance for 11–14-year-olds, and delayed evidence of an effect on the school attendance rates of 15–16-year-olds. School attendance for elementary and middle school children is unaffected by high school entry, as expected.[26] We only see large increases in attendance for 17–18-year olds, exactly the population

25. We also show in Section 6 that our results are not dependent on our chosen set of fixed effects. Findings are similar if we exclude county-by-birth-year fixed effects or high school year-by-birth-year fixed effects.

26. High school entry could affect attendance at earlier ages if it changes the perceived value of finishing earlier school. But we expect this impact to be smaller than the direct effect for older students.

that we think should respond most to high school entry.[27] While common schools and emerging "middle schools" could enroll 13–16-year-olds in the absence of a high school, it was less likely that older students would attend those schools.

Effect sizes grow rapidly, both in absolute and percentage terms, for high school-aged students with access to schools. These growing estimates are consistent with past work on the high school movement, which showed the largest attendance effects in later years as schools expanded enrollment and more high schools were built. Because we focus on high schools built before 1909, our median construction date is 1885. Goldin (1998) shows that the high school movement expanded most dramatically from 1910 onward, peaking around 1940. This timing lines up with our estimated attendance effects. Places that initially established high schools were also places that later invested more in them and saw higher attendance than places without a high school nearby.

Our estimates are large: being near a high school increases the probability of attending school by about 10 percentage points for 17 and 18-year-olds, or a 28 percent increase relative to 1870 attendance rates. This percentage change likely understates the actual percent increase in high school attendance. Around 30 percent of children aged 17–18 reported any school attendance in the previous year in 1870, but we know that very few of them were graduating high school – fewer than 5 percent did by 1880, according to aggregate records (NCES 1993). As discussed in Section 3, this discrepancy is due to the census question: enumerators were asked to report whether a child attended school during the year. Many may have attended junior high schools, church Sunday schools, or similar facilities part-time, but we would not consider them to have attended high school. Using the aggregate graduation figure as a baseline probability would imply a much larger first-stage increase in schooling, but would overstate the impact since some students encouraged to attend high school did not graduate. The actual change in attendance rates is likely between that number and the baseline probability calculated from reported census attendance.

Figure A4 (Panel A) shows that attendance effects are almost entirely driven by an increase in

27. Given the timing of the censuses, many children who report being 18 when it is taken were of high school age during the period covered by the attendance questions.

'full-time' attendance: children who reported attending school in the prior year but did not report an occupation in the census. Figure A4 (Panel B) shows a small positive (age 17-18) or slightly negative effect (other ages) on the probability of reporting attendance but also having an occupation ('part-time' attendance).[28] Attending high school was a more demanding form of education than what otherwise might be available to local youth and required more time.

Figure 3 shows 17–18 year-old attendance event study estimates separately for men and women. Attendance for men and women responds similarly in the immediate years after a school opens. However, on a percentage-point basis, the male attendance effect is larger in the later years of our sample. We summarize the average impacts shown in these event studies using a pre-post specification and report the coefficients in Table A4. We report both the overall pre-post comparison coefficients and a truncated sample window comparing children aged 0–14 to those aged 18–32 when high schools opened. This truncated comparison reveals that women responded faster than men. And while the coefficients for longer-run effects are similar, the growth for women was larger on a percentage basis.

4.3 Adult labor market outcomes

To estimate the effects of high school access on adult outcomes, we estimate analogous models:

$$AdultOutcome_{ictk} = \sum_{a=-30}^{60} \beta_a (Treated_i \times \mathbf{1}[\text{age} = a]_{ictk}) + \lambda_b + \delta_c + \eta_t + \gamma_{ob} + \phi_{kb} + \varepsilon_{ictk} \qquad (2)$$

The key difference from Equation 1 is that the outcome variable is now a labor market or demographic characteristic measured in adulthood. Another difference is that the sample now focuses on children from age 10–18 whom we observe in a childhood census record.[29] If we see a 13-year-

28. These proxies are as defined in Goldin and Katz (2011). As they note, they are imperfect measures: some children could work part-time or summer jobs and still attend school full time.

29. Our sample is limited to children that we can link to their age 20s or 30s outcomes. We exclude 19 and 20-year-olds since their residential locations at those ages might be endogenous to high school openings. We show results using only the age 17 and 18 sample to do long-run linking in Section 6.

old in the census, we will never be able to see their high school attendance, since in the next census they will be 23. But they might have been affected by the high school, so we can calculate their age at the time of high school exposure and see whether their adult labor market and demographic outcomes at ages 23 and 33 change.

Equation 2 follows the same empirical strategy as our attendance equation and has the same interpretation: How did having a high school within 0.5 miles relative to 3 or more miles away affect outcomes in the years leading up to and following high school constructions? Pre-trends are interpreted similarly. If you were a woman who grew up in a town 10 years before it got a high school as opposed to 20 years before it got a high school, were you more likely to enter the labor force in your 20s or 30s? In other words, were outcomes of children who grew up in these towns already shifting, or does it appear that any shift in long-run fortunes corresponds to the opening of the school?

Labor force participation

Figure 4 shows the results of our event study specification, with labor force participation as the outcome variable. The x-axis of these graphs shows ages relative to the opening of a high school. For this outcome (and all subsequent young adult outcomes), we present four specifications: results for men and women at ages 20-28 and 30-38.

Figure 4 shows modest increases in the probability of labor force participation in the overall sample. In the immediate aftermath of high schools opening, we see an average increase in labor force participation of approximately two percentage points for 20–28-year-olds and 1.5–2 percentage points for 30–38-year-olds. But these aggregate effects mask important heterogeneity by gender. Male labor force participation in the age-20s sample declines and there is a precisely estimated null effect of high school access on labor force participation for men in their 30s. The small, negative result for men is, at first glance, a puzzle. However, men in their 20s might be more likely to attend college after high school entry, which would delay labor force participation. Consistent with this explanation, when we limit our sample to 27 and 28 year olds, we observe no effect on labor force participation (Figure A22). We also find that high schools increase the probability of men reporting school attendance at age 20–28 in a similar magnitude as the decrease in LFP (Figure A5).[30]

For women, we see large and growing effects on labor force participation. Women aged 20-28 with childhood high school access are immediately more likely to be in the workforce. For women age 9–11 when high schools opened, we estimate an approximate 3.5 percentage-point increase in labor force participation. This is a 13 percent increase relative to the 1880 mean labor force participation rate for women in our linked sample. Effect sizes for women grow over time, consistent with the first-stage effects on attendance reported in Figure 3. Interestingly, the labor force effect attenuates by roughly a third for women in the 30–38-year-old sample. This decrease suggests that initial boosts in labor force participation are not fully sustained as women advance in their careers; we discuss this pattern in more detail as we examine other labor market and demographic outcomes.

Combining the female and male labor force results indicates that high schools were an important contributor to the convergence in labor force participation observed in aggregate data during the first half of the 20th century. We use our estimates to more rigorously quantify this contribution in Section 7.

Job choice and occupational income

High school access reshaped the types of jobs that young adults entered, even for groups whose labor force participation remained unchanged. We estimate our main specification with a series of indicators for whether a young adult worked in a specific occupation. We begin by focusing on clerks, an occupation that grew during this period and often required writing and arithmetic skills taught in high schools. We rely on the IPUMS occupation “clerical” category, which includes bookkeepers, secretaries, and stenographers. These were among the most common skilled occupations that women began entering in large numbers during this period.

30. Men and women also retrospectively report higher college attendance in the 1940 Census for cohorts affected by high school openings (Figure 12).

In Panel A of Figure 5, we show results with an indicator for clerical work as the outcome. These figures show that the probability of being a clerk increased for affected men and women after high schools opened. The magnitude of the effect is larger for women than men—we see a 4 percentage point increase in the probability of reporting a clerk occupation for 20–28 year old women two decades out, relative to an approximate 2 percentage point increase for men in their 20s. For women, this was a large percentage increase, given that only 0.1 percent of 20–28-year-old women reported working as a clerk in 1860. This shift toward white-collar labor supports the argument that high schools imparted valuable skills (e.g., in mathematics and typing) to their students, who then turned those skills into clerical jobs. Consistent with growing attendance, the effects increase for cohorts born in the years after a high school opens. Effects for both men and women decline in magnitude for age 30 outcomes, particularly for women. This implies that high schools were particularly important for first or early jobs in newly growing occupations.

Next, we test whether men and women were more likely to enter "professional" careers after gaining access to high schools. This IPUMS occupation category includes doctors, professors, scientists, and engineers. Panel B of Figure 5 shows increases in the probability of being employed in these professions for cohorts with access to a high school. Interestingly, we see similar percentage-point effects for men and women on this outcome. However, the increase for women is a much larger percent change, since only 1.7 percent of 30–38-year-old women in 1880 worked in a professional occupation, compared to 4.2 percent for men. For women, these results are driven by shifts into teaching occupations (Panel B of Figure A9), one of the few professional occupations available to women at this time.

Figure 6 shows a summary of the main occupational results for men and women in the age 30s sample. This figure plots the coefficients from a pre-post stacked difference-in-differences specification rather than using relative timing bins. The coefficients capture the average effects of having access to high schools in the years after they open. Figure 6 shows that occupational reallocation from lower prestige and pay occupations toward clerical and white-collar industries occurred after high school entry for both men and women. However, the reallocation effect appears larger for men. In contrast, the labor force participation effect is larger for women, consistent with the results shown in the event study figures.[31]

To further summarize the combined effects of high schools on labor force participation and occupational choices, Panel A of Figure 7 shows results with logged imputed occupational income as the outcome variable.[32] For this analysis, we use IPUMS's 1950 occupational scores, which assign the median 1950 total income to similar occupations in earlier periods. Occupational income increases by 15-20% for 20–28-year-old women in the cohorts with high school access, with little impact on men due to offsetting decreases in labor force participation and increases in occupational prestige. Panel A of Figure A10 shows that while the occupational income result for men increases from their 20s to 30s as the labor force participation effect disappears, the effects for women attenuate. This result is consistent with the similar attenuation observed in women's labor force participation and occupational choice results.

For context, Figure 7 also shows the effects of high school access on sample children's parents' log occupation score and labor force participation. We measure parent outcomes when we observe the children at age 10-18. Parental outcomes are not a valid placebo test in this context because they could be affected by high schools. There may be spillover effects from access to high school that affect everyone in a town, even if they did not attend a high school. More directly, some of the children treated by a high school themselves will be parents later in our sample window. That said, Figure 7 shows that, as expected, the children themselves are the most affected by high school, and we do not see widespread evidence of differential parental sorting in the years leading up to and following high school entry that could explain our results.[33]

31. Figure A6 shows the summary figure for outcomes measured at ages 20-28. Figures A7-A9 show the event study results for additional occupational outcomes.

32. Actual income is only available in the 1940 census.

33. In Section 6, we show results for children that condition on parental labor force participation and log occupational income. For the reasons discussed above, this is likely controlling for part of the impact of high schools, but we show that it does not severely attenuate our results. Figure A10 shows that parental outcomes are closer to the magnitude of age 30 outcomes, unsurprisingly given that those results generally attenuate relative to age 20 impacts.

5 Heterogeneity and mechanisms

In this section, we describe possible mechanisms and intermediate effects that link our first-stage increases in high school attendance to our downstream labor market outcomes. We pay particular attention to the heterogeneity that we observe by gender. We also explore heterogeneity by family occupational income, which helps us explain why and for whom high schools were most effective.

5.1 Migration

High schools prepared students for jobs that might be available in greater numbers only outside their hometowns. For example, while clerical work was needed nationwide, demand for these jobs grew particularly rapidly in urban areas. Geographic migration is an important intermediate outcome that links high school access to job opportunities.

We proxy for geographic mobility by measuring the distance between a person's childhood home and the place they are observed in the census during their 20s and 30s. We estimate the probability of moving at least 50 or 100 miles from a childhood home. Figure 8 shows that both men and women are more likely to be geographically mobile if they have high school access. Effect sizes are similar for both men and women. During ages 20-28, both are about 2.5 percentage points (or 9 percent) more likely to be 50 miles away from their location a decade earlier. And unlike many of the labor market results in Section 4, the geographic mobility effect size increases when we look at outcomes for age 30-38. High school access encouraged a geographic move toward jobs that rewarded a high school education.

5.2 Marriage and fertility

The mobility results in the previous subsection help explain how high schools allowed children to access jobs that required additional human capital. But these results cannot explain the pattern of heterogeneity across men, women, and age groups that we observe in labor market outcomes.

Our labor force participation and occupational choice results imply that high school induced young adults, especially young women, to enter higher-paying careers. But why was much of this growth for women lost in their 30s?

Our estimates in Panel A of Figure 9 show evidence that the share of married people aged 20–28 decreases. We see little short-run change for men; women drive the delay. By age 30–38, it appears that the delay is mostly over — while there is a slight decrease in the probability of marriage for women at that age, it is difficult to distinguish from a mild pre-trend in the years leading up to high-school entry. Panel B of Figure 9 shows similar patterns in fertility, measured as the probability of a child being in the same household as a respondent. Fertility falls for women in their 20s by a similar magnitude as marriage, but this effect also attenuates by age 40.

These marriage and fertility findings are consistent with the attenuated occupational and income results for women that we observed in their age 30 outcomes. While high school access appears to have led to a lasting earnings boost for some women, other married women left the labor force in their 30s. This reflects existing gender norms and workplace cultures for these cohorts of women. "Marriage bars," the practice of not hiring married women (or of firing women who married), were common in clerical, teaching, and other fields that saw a disproportionate increase in women's participation in the early 1900s (Goldin 1988).[34]

1940s labor market and schooling outcomes

Our preferred strategy is to measure long-run outcomes by projecting 10 and 20 years forward from childhood. But for a subset of our sample, we can link children to the 1940 census, which includes labor-market and educational outcomes not available in other census years.

Figure 12 shows these results for both men and women. We estimate specifications for the baseline model and additional models that condition on high school-by-birth-year fixed effects. Results are noisier than our preferred specifications because this exercise relies on longer links that are difficult to make: our median treatment date is in the late 1800s, so many of these children

34. Tables A5 and A6 summarize all of our main long run results for young adult women and men, respectively.

need to be linked 30 or more years forward to 1940. That said, Panels A-D show results that are broadly consistent with our shorter-run analysis. High schools appear to boost long-run labor supply for women, but not for men. Access makes both men and women more likely to enter clerical occupations, with larger effects for women, and it makes both men and women more likely to enter professional occupations. These extensive and intensive margin results translate into sizable predicted occupational score gains (Panel D), particularly for women.

Panels E and F show outcomes available only in the 1940 census. High schools appear to increase the probability that both men and women will retrospectively report attending college by about 5 percentage points in the long run. High schools also increase reported wage income (Panel F). The 1940 income data can be difficult to interpret, since they does not include non-wage income. Non-wage income was a sizable portion of compensation for both occupations we expect high schools to have shifted people out of (farming) and into (business ownership). However, the results are directionally consistent with the large increases in more prestigious, well-paying jobs (e.g., clerks) observed in the occupation data.

Family background and income

Figure 10 shows results split by family characteristics of the children in our sample. We estimate attendance impacts separately for children of US- and foreign-born mothers and after splitting the sample at the median father occupational score, with an additional category for children with no father present in the household.

The most striking pattern is the large gap in attendance responses by paternal income. Children from higher-income families are roughly twice as responsive to a nearby high school opening as children from lower-income families. For boys with high-income fathers, the attendance treatment effect is approximately 7-8 percentage points, compared to roughly three percentage points for boys from lower-income households; the same pattern holds for girls. While public high schools were "free" to attend, high school attendance required a baseline level of academic preparation more commonly present in wealthier households. Moreover, lower-income families faced a steeper opportunity cost: a teenager's labor represented a meaningful loss of household income at a time when child labor was common.

Figures A11-A12 show that this attendance heterogeneity is mirrored in many of the longer-run outcomes we study. Among women, the labor force participation increase, the shift into professional and clerical occupations, and the occupational income increase are all larger for those from wealthier families. Men follow a similar pattern, though the differences are smaller. Occupational shifting into higher-paying clerical and professional jobs is larger for men from high-income households. While the early high school construction boom benefited people from both richer and poorer households, children from richer households were best positioned to take advantage of new opportunities from increased educational access.

Black children

We next examine the effect of high school access on Black children. As discussed in Section 2, we have strong *ex ante* reasons to believe that most Black students during this period would rarely benefit from a nearby high school. This is partly because the existing common school structure was much weaker in the South, where most Black residents lived; it is difficult to benefit from a high school if you did not complete earlier (7th and 8th) grades. Moreover, almost all schools built in the South—where the vast majority of Black individuals lived in the pre-Great Migration period—were segregated, and very few communities provided a separate Black school. We are underpowered to detect effects for just these locations. We instead run our baseline regression with all schools using the subsample of Black children to measure outcomes. On the one hand, this approach may understate the actual transformative impact of the few Black high schools during this period. On the other hand, it highlights that, in *aggregate*, high schools built during this period had a negligible effect on the Black population.

Figure 11 shows the estimates for the Black subsample. Panel A shows a precise null effect on attendance for Black children, consistent with the historical record. We also see no downstream impacts on any of the long-run outcomes we study, including labor supply, occupational choice, and demographic and geographic mobility. We observe an imprecise increase in the probability of landing a professional job many years after a high school opens, but this may be due to general equilibrium and spillover effects. We cannot rule out small effects of high schools on these long-run outcomes. But the general pattern of results suggests that, in aggregate, Black children were not meaningfully affected by these high schools.

These null results highlight that the pre-1909 high school construction boom exacerbated existing white-Black human capital gaps. Access to a high school was transformative for the white population, particularly for women. But because it did not affect the Black population, at the same time that white women closed gaps with white men, Black children fell further behind. This gap later became the impetus for programs like the Rosenwald Schools in the 1910s and 1920s, which built thousands of schools in the South for Black children and had significant long-term effects (Aaronson and Mazumder 2011).

These findings also help confirm the appropriateness of our estimation strategy. The historical record suggests that Black children did not attend high schools in any large numbers during this period. While high schools likely had spillover effects on those who did not attend (see Section 7), it would be surprising to find large effects for Black children during this period. Had we found them, it would have suggested that a city-specific, time-varying confounder correlated with school construction, rather than high schools themselves, drove our results. Instead, our results are consistent with the idea that early high schools primarily affected whites, widening racial gaps.

6 Robustness

Our baseline estimation uses a treatment definition of being in a city within 0.5 miles of a high school-containing city and a control group definition of being at least 3 miles away from a city with a high school. This estimates a specific effect of being immediately proximate to a high school, but other thresholds are reasonable choices and allow us to trace the relationship between high school access and geographic distance in more detail.

Figure A13 shows that our results are robust to alternative distance thresholds for both our treatment and control groups. We show results where we hold the current treatment group definition fixed at 0.5 miles and adjust the control group radius to 0.5, 1, 2, 4, or 5 miles. Similarly, we hold the current control group threshold fixed at 3 miles and expand the treatment radius to 1, 2, or 3 miles. Results across all cases show a positive, statistically significant effect on attendance, and the treatment effect moves in the expected direction. All else equal, as we expand the treatment radius, effect sizes shrink because we are including people farther from the high school in our treatment group. As we expand the control group threshold, effect sizes increase because we are comparing people close to a high school with those farther away. Figures A14 (women) and A15 (men) repeat this exercise for adult outcomes. Consistent with the attendance results, our qualitative findings do not depend on specific control and treatment group definitions, but we observe larger impacts from high schools as we tighten the treatment radius or widen the control radius.

To verify that our results are driven by actual proximity to high schools and not spurious specification artifacts, we conduct a test using randomized distances. We replace each city's actual distance to the nearest high school with a distance drawn without replacement from the distribution of actual miles to the nearest high school in our sample. This results in a sample with the same number of treatment and control cities as in our baseline analysis, but with treatment status randomly distributed. We then re-estimate our baseline specifications for attendance and adult outcomes using the randomized distances. The top panel of Figure A16 shows that for both men and women, we estimate precise null attendance effects using these randomized distances; both sets of results are distinct from our attendance effects using real distances (bottom panel). Figures A17 (women) and A18 (men) repeat this exercise for adult outcomes. We consistently find null results when using a randomized distance.

Next, we perform a series of additional specification and sample tests on our baseline sample. All these tests are shown in pre-post summary form in Figures A20 (attendance), A21 (long run, women), and A22 (long run, men). We also estimate each event study corresponding to these robustness checks and show the results separately by outcome in Figures A23–A33.

First, we adjust our sample ages. We expand the ages for our attendance results to 14–18. We also limit the sample of children that we link to our long run outcomes to age 17–18. The attendance results show that we obtain attenuated but still positive and statistically significant attendance impacts when we expand the sample. For the long-run results, the restriction to ages 17–18 shows that when we use the same baseline sample as our main attendance results, we estimate similar, though less precisely estimated, results.

Second, we present results for a subset of 10–18-year-olds whom we can link to their childhood locations at ages 0-8. We condition on respondents who have moved less than 5 miles between the ages 0-8 and 10-18. This helps address a concern about selective migration: our effects could be driven by families who moved to an area *because* it had a high school. If that migration also correlates with other downstream outcomes, we could mistakenly attribute some of our results to it. We see noisy but similar patterns of results using the subset of non-moving households.

Third, we present results that condition on mother's and father's occupation scores, separate indicators for mother's and father's labor force participation, and an indicator for having railroad access. This helps address a similar differential migration story as the one above, as well as a potential concern that something else about these cities is changing at exactly the same time as high schools enter.[35] Figure A19 shows that while cities that received the high schools were positively selected on railroad access, this selection is smooth across the exact treatment timing of high school openings. This suggests that railroad access is unlikely to explain our results. Consistent with this, we estimate similar results after conditioning on parental characteristics and railroad access.

Fourth, we estimate models limited to cities where we observe at least 20 or 25 age "bins" relative to high school construction. This eliminates a number of cities with unbalanced pre- or post-periods due to city incorporation right before high school formation, city mergers, and other

35. As discussed in Section 4, parental characteristics "overcontrol" in this context if they themselves are affected by the treatment.

similar circumstances. Results are similar, though in some cases less precisely estimated, when we make the panel more balanced.

Next, we estimate models with different sets of fixed effects. Our baseline specification includes county-by-birth-year effects, in addition to high school cohort-by-birth-year effects. We estimate models that replace county-by-birth-year with state-by-birth-year effects and obtain similar results. We also estimate a demanding specification that includes high school-by-birth-year fixed effects. This constrains comparisons to be among children who were different distances away from the *same* high school when it opened.[36] Results are similar across these specification choices.

Lastly, Figures A20–A22 show results with two-way clustering by nearest high school and county and two-way clustering by nearest high school and state. This check allows for broader geographic correlation of errors than our baseline model, which clusters by nearest high school. We obtain similar significance levels across all outcomes with these alternative clustering approaches.

Next, we show results with an alternative control group. Our baseline analysis uses cities that were more than 3 miles from a high school when it opened as controls. One potential concern with this approach is that these cities are fundamentally different from cities that received high schools and could be on a differential post-HS trajectory. As a robustness exercise, we use cities that *later* received a high school as controls for cities that built one earlier. We offset dates by 21 years: for example, a city that built a high school in 1900 is used as a control for a city that built one in 1879. In this case, we define 1879 as the focal year and calculate treatment timing relative to it for the control group. We then estimate versions of Equations 1 and 2 that trace out the relative differences in outcomes between later and earlier cities around the years when the earlier city built a high school, conditional on place-by-focal-cohort, year-by-focal-cohort, and birth-year-by-focal-cohort fixed effects. Figures A34-A36 show the results of this exercise, respectively for enrollment,

36. The baseline high school cohort-by-birth-year effects do not make this restriction. They allow pooled comparisons across high schools that were founded in the same year.

age 20s outcomes, and age 30s outcomes, overlaid with our baseline results.[37] Over the periods in which both methods yield pre-period estimates and treatment effects, we find a similar pattern of results.

In addition to these sample and specification checks, we check that our long-run results are not sensitive to the linking methodology that we use. Our default approach uses all links available in the Census Tree (Buckles et al. 2025). Figure A37 shows our baseline results alongside two additional linked samples. The Census Tree crosswalks provide information on which methods are used to establish links; we show results only for links where at least two methods agree. Next, we present results based on the Family Tree, a subset of high-quality Census Tree links derived directly from genealogical data. Figure A37 shows that our results are similar across all outcomes for both women (Panel A) and men (Panel B) across these linking methods.

7 Discussion and broader trends

7.1 Effects on aggregate female labor force participation

Our estimates provide the first opportunity to quantify the extent to which the dramatic increase in women's labor force participation during the late nineteenth and early twentieth centuries can be attributed to the high school movement. In this section, we perform back-of-the-envelope calculations that link our micro-level treatment effects to aggregate trends in female employment.

According to Goldin (2006), labor force participation among women aged 25–44 rose from approximately 18 percent in 1890 to 35 percent by 1940. For younger women (those in their twenties, the focus of our analysis), employment rates were even higher and rose more rapidly. Census data indicate that labor force participation for women aged 20–28 increased from roughly 20 percent in 1870 to approximately 37 percent by 1930, a 17 percentage point (85%) increase

37. The estimation window for this analysis is shorter than the baseline because we only use early-treated units in the treatment group (since we need the later-treated units as controls) and because the post-period needs to be truncated so that we avoid picking up a treatment effect for the control units as they build their own high schools.

in the share of young women engaged in formal work. As we discussed above, access to high school expanded at a similarly rapid pace. High school graduation rates among 18-year-olds rose from approximately 9 percent in 1910 to over 50 percent by 1940 (Goldin and Katz 2008). The number of towns and cities with high schools expanded from fewer than 3,000 in the late 1800s to more than 20,000 by the early 1950s, a quintupling of access that far outpaced the growth in the school-age population over the same period.

To link the increase in female labor force participation to the expansion of high school access, we must first determine the share of the relevant population that was effectively "treated" by the high school movement. Our main empirical strategy defines treatment as living within 0.5 miles of a town or city with a high school. Using our geocoded panel, we estimate that by 1910, approximately 45% of the non-urban white population lived within easy commuting distance of a high school, up from roughly 15% in 1880. By 1930, this share had risen to approximately 70%. For a cohort of women reaching adulthood in 1920 (born around 1900), we estimate that roughly 55% had access to a high school during their teenage years, compared to approximately 25% for the cohort reaching adulthood in 1890. This implies that roughly 30 percentage points more women were "treated" by high school access from the 1860 to 1910 birth cohorts.

Our estimates indicate that high school access increased labor force participation among women aged 20–28 by approximately 4.9 percentage points (Figure 4 and Table A5). This effect represents an 18 percent increase relative to the 1880 baseline participation rate of roughly 27 percent in our linked sample. Importantly, this estimate captures the average treatment effect for women living near newly constructed high schools, compared with otherwise similar women in nearby communities without such schools.

To translate our person-level estimates into an aggregate effect, we combine our estimate of how women's labor force participation responds to local high school access with the increase over time in the share of young women who had access to high schools during their teenage years. Using our panel of places and high schools, we estimate that the fraction of the non-urban white population living within easy commuting distance of a high school rose from roughly 15 percent in 1880 to roughly 70 percent by 1930—a 55 percentage point increase in exposure. Applying our preferred reduced-form estimate for women in their twenties—an average post-opening increase in labor force participation on the order of a few percentage points—implies that initial high school construction mechanically accounts for a nontrivial share of the 17 percentage point rise in labor force participation for women aged 20–28 between 1870 and 1930:

$$\text{Proportion of female LFP from high schools} = \frac{0.049 \times 0.55}{0.17} = 15.9\%$$

However, this calculation understates the high school movement's true contribution to female labor force participation for several reasons. First, our treatment effect is estimated using variation in proximity to high schools, which captures the intensive margin of access but assumes that the high school movement has no effect on children living farther away from the new high schools. Second, our estimates focus on high schools built before 1909, capturing the ”first wave” of construction. The continued expansion of high schools through 1940, when high school graduation rates reached 50 percent, implies additional effects not fully captured in our analysis. In particular, towns and cities met the increasing demand for high school access by constructing new high schools and expanding existing ones to accommodate more students.

Figure 3 shows that attendance effects for 17–18-year-olds increased from approximately three percentage points immediately after school opening to nearly 10 percentage points two decades later. Similarly, labor force participation effects (Figure 4) exhibit dynamic patterns that suggest larger benefits as communities invest more in their schools. To estimate the long-run impact of this increase in attendance, we can map the average relationship between the attendance response and the labor force participation response and apply it to our estimates. Using our pre-post summary estimates (Tables A4 and A5), the observed ratio of labor force to attendance responses is roughly 0.94 (0.0494/0.0524). Assuming this result holds at the end of the panel, the 10 percentage point long-run female result that we observe at the end of our estimation panel (Figure 4) implies a 9.4 percentage-point increase in female labor force participation.

Using the same logic as above, we arrive at a new estimate of:

$$\text{Proportion of female LFP from high schools} = \frac{0.094 \times 0.55}{0.17} = 30.4\%$$

This back-of-the-envelope calculation suggests that the U.S. high school movement accounted for just under one-third of the increase in female labor force participation in the first half of the 20th century.

7.2 Effects on inequality

The previous subsection argues that the high school movement was responsible for a substantial proportion of the increase in the female labor force participation rate. In this section, we extend the same back-of-the-envelope framework to ask a complementary question: how did the high school movement affect existing gaps in school attendance and adult labor market outcomes across gender, race, and family income? The results show that while high schools narrowed the gender gap in labor force participation and occupational quality, they simultaneously widened Black-white gaps and reinforced class-based disparities.

We apply the same accounting logic as the previous subsection. The aggregate contribution of the high school movement to any outcome gap equals the differential treatment effect across groups, scaled by the change in the share of the population with high school access. We combine our treatment effects from Tables A4–A6 and Figures 10–11 with the estimated 55 percentage point increase in the share of the non-urban white population living within commuting distance of a high school between 1870 and 1930 (from approximately 15% to 70%).

Gender gaps

The gender gap in labor force participation is the margin on which high schools had their most transformative effect. Table A5 shows that high school access increased women's labor force participation at ages 20–28 by around 4.94 percentage points from an 1860 baseline of 13.8 percent. Table A6 shows a corresponding decrease in men's labor force participation of 1.51 percentage points, from a baseline of 86.2 percent, consistent with some men extending their education into college (Figure 4). These effects push in the same direction to narrow the gender LFP gap by 3.55 percentage points.[38]

Historical data show that aggregate labor force participation for women aged 20–28 rose from roughly 20 percent in 1870 to 37 percent by 1930, while men's participation was approximately stable at around 90 percent. The observed narrowing of the gender LFP gap was therefore on approximately 17 percentage points. Our estimates imply that the high school movement accounted for roughly 3.55 of these 17 percentage points, or about 21 percent of the convergence in male–female labor force participation rates for young adults.

Black-white gaps

The heterogeneity results in Section 5 and Figure 11 reflect the starkest type of inequality generated by the high school movement. Figure 11 shows a precise null effect on school attendance for Black children, and similarly null effects on adult labor market outcomes. These results reflect the de jure and de facto segregation that excluded most Black children from the newly constructed high schools, particularly in the South, where the vast majority of the Black population resided before the Great Migration.

Because the estimated effects on white children are large and positive while those for Black children are approximately zero (Figure 11), the high school movement mechanically widened existing Black-white gaps in every outcome we study. For example, the average school attendance

38. The calculation for that overall effect is $(0.0494 - (-0.0151)) \times 0.55 = 3.55$.

effect for white 17–18-year-olds is approximately 5.7 percentage points (averaging across the male and female estimates in Table A4). The corresponding effect for Black children is indistinguishable from zero (Figure 11, Panel A). Scaling by the change in treatment exposure, we arrive at an estimated 3.1 percentage point widening of the Black-white school attendance gap.

High-income versus low-income family gaps

Figure 10 documents striking heterogeneity in attendance effects by parental income. Children from families with above-median father occupational scores are roughly twice as responsive to a nearby high school opening as children from families with below-median father occupational scores. Specifically, for boys with high-income fathers, the attendance treatment effect is approximately 7–8 percentage points, compared to roughly 3 percentage points for boys from lower-income households; the same pattern holds for girls. This 2:1 ratio implies that the high school movement widened class-based gaps in school attendance by 2.5 percentage points.[39]

This gap is economically meaningful: it implies that for every 100 children newly exposed to a high school, roughly 4–5 additional children from wealthy families would attend relative to children from poorer families. At scale, across thousands of newly constructed high schools, this differential response compounded existing inequalities in human capital accumulation. That said, Figure 10 shows that there *was* an enrollment response even among low-income children, unlike for Black children.

These calculations reveal a tension in understanding the effects of the high school movement. The same institution that meaningfully narrowed the gender gap in labor force participation and occupational quality simultaneously widened both racial and class-based gaps in school attendance and labor market outcomes.

39. The calculation to arrive at that overall effect is $(7.5-3.0)\times 0.55 = 2.5$.

7.3 Estimating aggregate effects

A key question is whether high schools affected only the cohorts who directly gained access and attended, or whether openings also changed outcomes for other people in the same local economy who did not (or could not) attend the newly constructed high schools. This matters for two reasons: first, our main estimates are relative effects; we compare high-school-aged individuals who were "treated" by proximity to individuals slightly farther away in the same county and birth cohort. If high school entry shifted local labor market conditions more broadly, some of those changes could affect both the treated and the control group, changing how our reduced-form estimates map into the overall effect of high school expansion on aggregate outcomes. Second, the high school expansion could plausibly generate either positive or negative spillovers, both of which have broader implications for our understanding of the value of education and modern-day arguments about signaling, sheepskin effects, and human capital. A conceptual argument for positive spillovers relies on beliefs that access to education produces important skills, higher levels of productivity, more business formation, and occupational upgrading that benefits (or at least, does not harm) nearby non-attendees. A conceptual argument for negative spillovers rests on the idea that credentialing of high school attendees reduced opportunities for workers without access, but did not affect aggregate local outcomes.

Our results include three complementary exercises that help assess the likely magnitude and sign of spillover effects: (1) effects of high schools on the parents of treated children: older cohorts who largely did not attend the new high schools; (2) outcomes among individuals living 3–5 miles away from high schools: a group of people with low attendance responses who worked in the same labor market as high school attendees; and (3) effects of high schools on Black children: a group with sharply limited access to schools and attendance rates in this period, who nonetheless operated in overlapping labor markets with high school attendees. Taken together, these analyses suggest that spillover effects exist but are modest in magnitude relative to the direct effects on the cohorts with access. In other words, we do not see patterns consistent with large negative displacement on people who did not attend high school, but operated in the same labor market as attendees.

Effects on parents

We first examine whether high school openings affected the labor market outcomes of parents of exposed cohorts. These adults generally were not eligible to attend the new high schools themselves, but they lived in the same places as ‘treated’ individuals and thus could be affected by changes in local labor demand or local occupational structure. Parental outcomes are not a placebo; they are plausibly affected through multiple channels, including both the labor market and household income effects from children, so “non-zero” effects are not evidence against identification.

Figure 7 directly compares event-study responses for daughters and sons (who could attend) to responses for fathers and mothers (who largely could not) in the same households, focusing on labor force participation and occupational income in the child’s early adult years. The key empirical pattern is that the post-opening changes for daughters are large and sustained—especially for labor force participation and log occupational score, while the corresponding effects on parents are much smaller in absolute magnitude (Figure 7). In labor force participation, mothers show a small positive shift after high school entry, while fathers’ labor force participation is slightly negative; both are small compared with the daughter series. In occupational income, both parents move modestly upward relative to the very large increase for daughters (Figure 7). These comparisons are informative because they hold fixed the same local environment and timing, and they speak directly to equilibrium channels that would affect adults in the area rather than school-age cohorts.

A closely related comparison for outcomes measured in individuals’ 30s yields the same broad conclusion. In Figure A10, daughters’ labor force participation and logged occupational income rise substantially after high school entry, while the equivalent parental series again exhibit much smaller movements.[40] The fact that parental outcomes move at all is consistent with some combination of household responses and local labor market spillovers. However, the difference in

40. Table A7 summarizes these results.

magnitudes—large gains for the directly exposed cohorts versus modest changes for parents—suggests that the key channel in our main estimates is not a broad-based labor-market shock affecting all nearby adults equally, but rather access-driven changes concentrated among school-age cohorts.

Effects on adjacent places

A second approach to estimating spillover effects uses distance to separate the effects of access and attendance from general local labor market exposure. In our main empirical design, treated individuals live within 0.5 miles of a new high school and control individuals live at least 3 miles away. To investigate whether high school entry affected people who were exposed to the same local labor market but were unlikely to attend the high school, we consider outcomes for individuals living 3-5 miles from a high school and compare them to even more distant control individuals. Key to this analysis is a reminder that our distance is measured such that everyone in a place lives in the same point in space, so individuals living in one town all reside at the same coordinate, and people 3–5 miles from them must be in adjacent towns and well-defined places.

The first-stage evidence supports the idea that this 3–5 mile ring of individuals experienced a much smaller direct attendance response than the baseline treated group. In Figure A38, the baseline treatment definition generates large and rising attendance effects for both women and men, while the 3–5 mile treatment definition leads to flat attendance effects, close to zero in magnitude. This is precisely the pattern we expect if close proximity to a high school within 0.5 miles captures direct access/attendance, whereas those individuals living 3–5 miles away capture a group with limited attendance but potential exposure to the local labor market.

Turning to adult outcomes, Figures A39 and A40 compare the baseline treatment definition to the 3–5 mile treatment definition for age-20 and age-30 outcomes, respectively. At age 20, the 3–5 mile series shows effects close to zero across major outcomes, especially labor force participation and occupational income, while the baseline treated group shows the large positive effects for women and the different pattern for men that we highlight in the main results. This contrast is informative because it shows that the substantial early-adult gains we estimate for women are tightly linked to "direct access" to a high school, and do not arise mechanically from a broad labor market shock that affected everyone in a wide surrounding area equally.

By age 30, however, the 3–5 mile treated group for women shows more evidence of positive effects on some outcomes. In Figure A40 women in the 3-5 mile ring exhibit a positive change in labor force participation that is smaller than, but in the same direction as, the baseline group of treated women. Similarly, Figure A40 shows positive movements in logged occupational income for women in the 3-5 mile ring that are again smaller than the baseline treated effect, but still clearly present. In contrast, the 3–5 mile series for men tends to remain near zero in these panels. This pattern is consistent with modest positive spillovers that are detectable for women by age 30, while not pointing to large positive or negative spillovers for men.

Two interpretive points are crucial here. First, because the 3–5 mile ring does not have a perfectly zero attendance response (Figure A39), these estimates are best viewed as an upper bound rather than a clean estimate: a small amount of direct attendance at these distances could contribute to the observed adult outcomes. Second, the fact that the 3–5 mile outcomes are generally much smaller than baseline treated outcomes—especially at age 20 (Figure A39)—makes it difficult to reconcile our main estimates with a story in which nearby residents who did not attend experienced large labor market harm via displacement. If high schools had substantially cannibalized clerical and professional opportunities from nearby non-attendees, we would expect clearer negative effects in the 3–5 mile ring on occupational outcomes and labor force participation, particularly in early adulthood. Instead, Figures A39 and A40 show effects that are mostly near zero or modestly positive, especially for women at older ages.

Effects on Black children

A third piece of evidence comes from the sample of Black children. In this period and setting, Black children were often prevented from attending, or were much less likely to attend, high schools. This means that any effect of nearby high school openings on the adult outcomes of Black children must operate through local labor market spillovers rather than through direct school attendance. Figure 11 shows that high school entry has essentially no detectable effect on Black attendance at ages 17–18, and the effects on adult outcomes for labor supply, clerical and professional employment, marriage, and geographic mobility are generally centered near zero with no sustained post-opening trend. This pattern is consistent with spillovers being limited in size, at least for this group.

At the same time, this evidence must be interpreted in the context of racial segregation and labor-market segmentation. If local labor markets were partially or entirely disjoint across race, then even economically meaningful spillovers of high schools to the labor markets with a concentration of white workers might not translate into measurable gains (or losses) in economic opportunities for Black workers. So, these null results do not, on their own, show the absence of spillovers; they do, however, provide a valuable check against the hypothesis that high school openings generated large, local, economy-wide shocks that predictably moved outcomes for all nearby residents regardless of high school access.

7.4 Implications for aggregate labor market effects

Section 7.1 uses our reduced-form estimates to quantify how much of the national rise in women's labor force participation could plausibly be attributed to high school expansion, combining our estimated effect of high school access on women's labor force participation (e.g., Figure 4 and the corresponding pre- and post- summaries) with the share of cohorts with high school access over time. This exercise is mechanically sensitive to spillovers because our reduced form is estimated relative to a control group that could itself be affected by nearby high schools if high schools affect labor markets more broadly.

Conceptually, if high school openings increased labor market opportunities for non-attendees in the same area (positive spillovers), then our treated-control comparison would understate the total effect of the high school movement on aggregate female labor force participation, because some gains accrue to individuals in the control group. If, instead, high school openings reallocated a fixed set of jobs toward newly credentialed cohorts and away from nearby non-attendees (negative spillovers), then our treated-control comparison could overstate aggregate gains. The three empirical checks above collectively point to either small or modest positive spillovers to groups that were not directly affected by high school access. These potential spillover effects are most visible for women by age 30.

Taken together, these results suggest that the aggregate contribution of high schools to the changing female labor force participation rate in Section 7 is unlikely to be dramatically overstated by negative spillover effects on nearby non-attendees. If anything, the pattern of modest positive movements among mothers and among women living 3–5 miles away by age 30 is more consistent with the possibility that our reduced-form estimates may be a lower bound on the total effect of high school expansion on aggregate female labor force participation. At the same time, because the evidence for spillovers is generally small in magnitude relative to the direct effects for the cohorts with access, we view the likely equilibrium adjustment to our aggregate accounting as modest, and we emphasize the Section 7.1 contribution numbers as an internally consistent reduced-form-based decomposition rather than a fully general-equilibrium estimate.

8 Conclusion

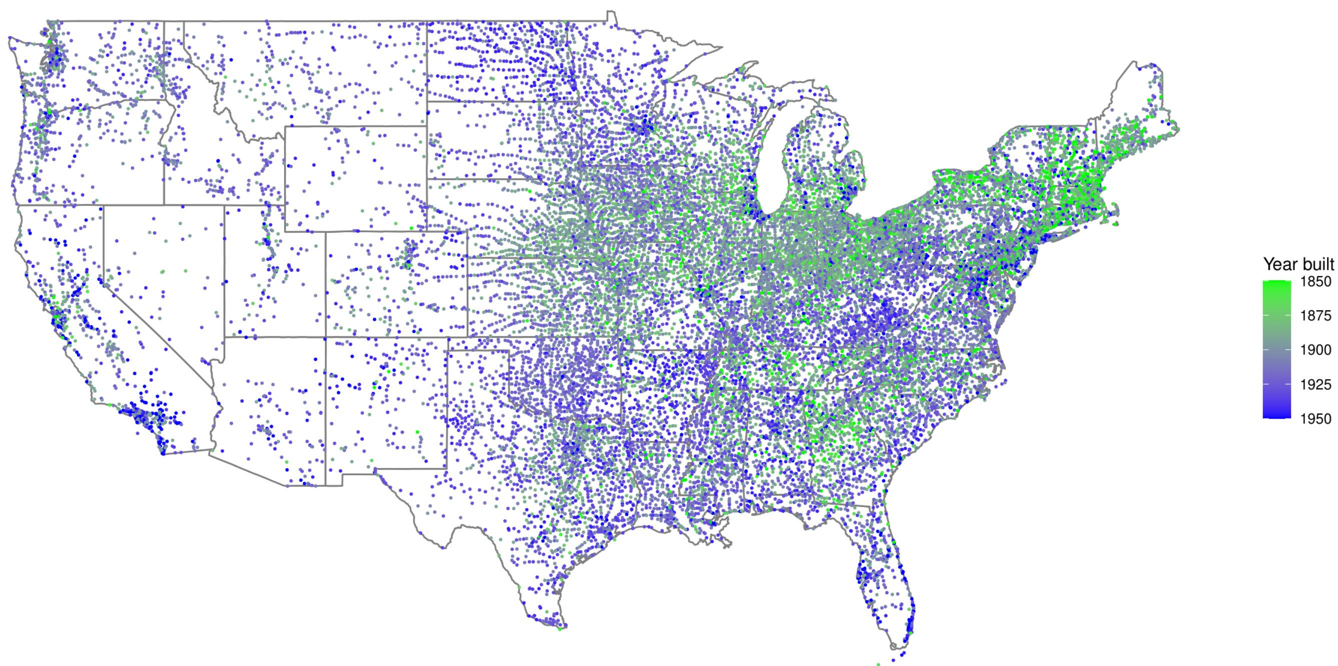
The rise of the high school was the defining transformation of the American educational system in the twentieth century. In this paper, we construct a new granular panel of U.S. high schools to provide the first causal estimates of how this construction boom reshaped the lives of the children who lived through it. Linking the sharp timing of local high school openings to individual-level census records, we document that the "first wave" of high school construction had dramatic effects on the U.S. economy.

We find that access to a local high school increased attendance among children aged 17–18, with effects that grew over time as communities deepened their investment in secondary education. This accumulation of human capital had downstream effects on the adult labor market, particularly for women. Access to high school explains approximately one-third of the aggregate increase in female labor force participation between 1870 and 1930. High schools drove occupational reallocation in the workforce, moving both men and women out of manual occupations and into the burgeoning clerical and professional sectors.

These economic shifts were accompanied by substantial demographic changes. For young women, the opportunity to attend high school and enter the white-collar workforce led to delays in marriage and childbearing in their twenties. However, the attenuation of our labor supply results for women in their thirties highlights the persistence of gender norms and marriage bars that constrained female careers during this era. While high schools opened doors for young women, the early twentieth-century labor market often closed them upon marriage.

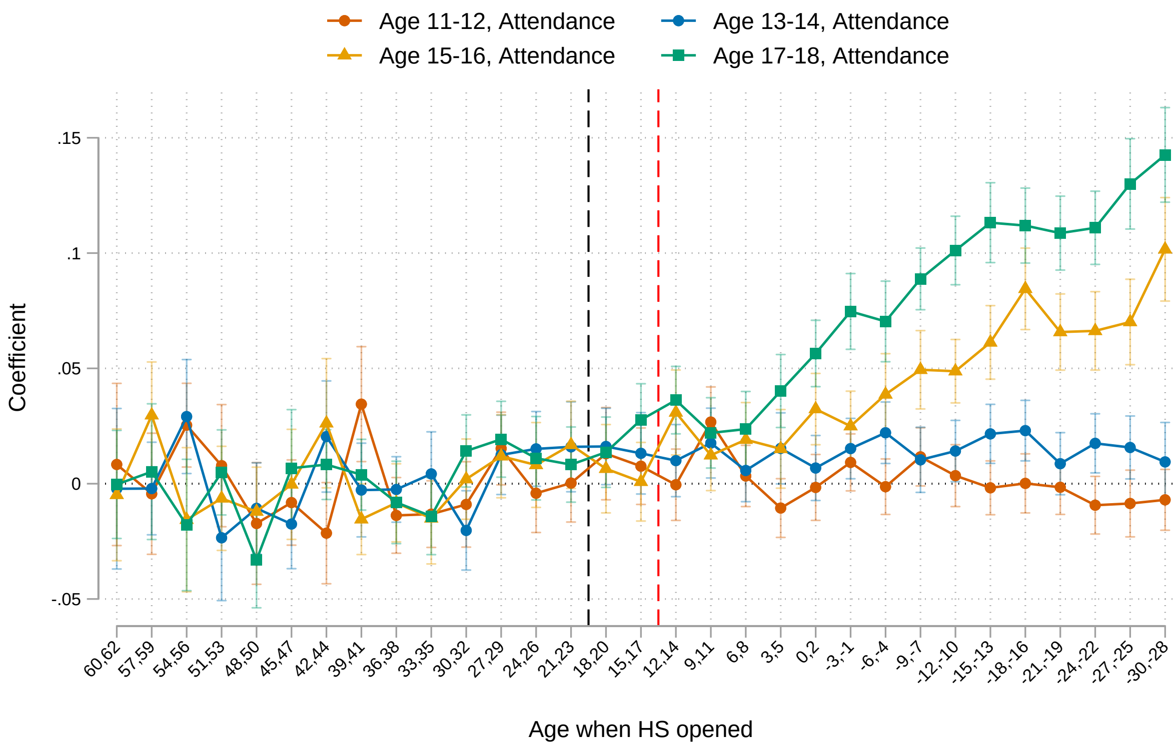
We show that the benefits of the high school movement were not widely shared across racial and class lines. We estimate precise null effects for Black children, consistent with the de jure and de facto segregation that characterized the era. And we estimate that while the high schools had positive impacts across the income distribution, effects were largest for children from wealthier families. Even as the early high school movement started to level the playing field between white men and women, it exacerbated existing human capital gaps.

Figure 1: Map of cities, by year of first high school constructed



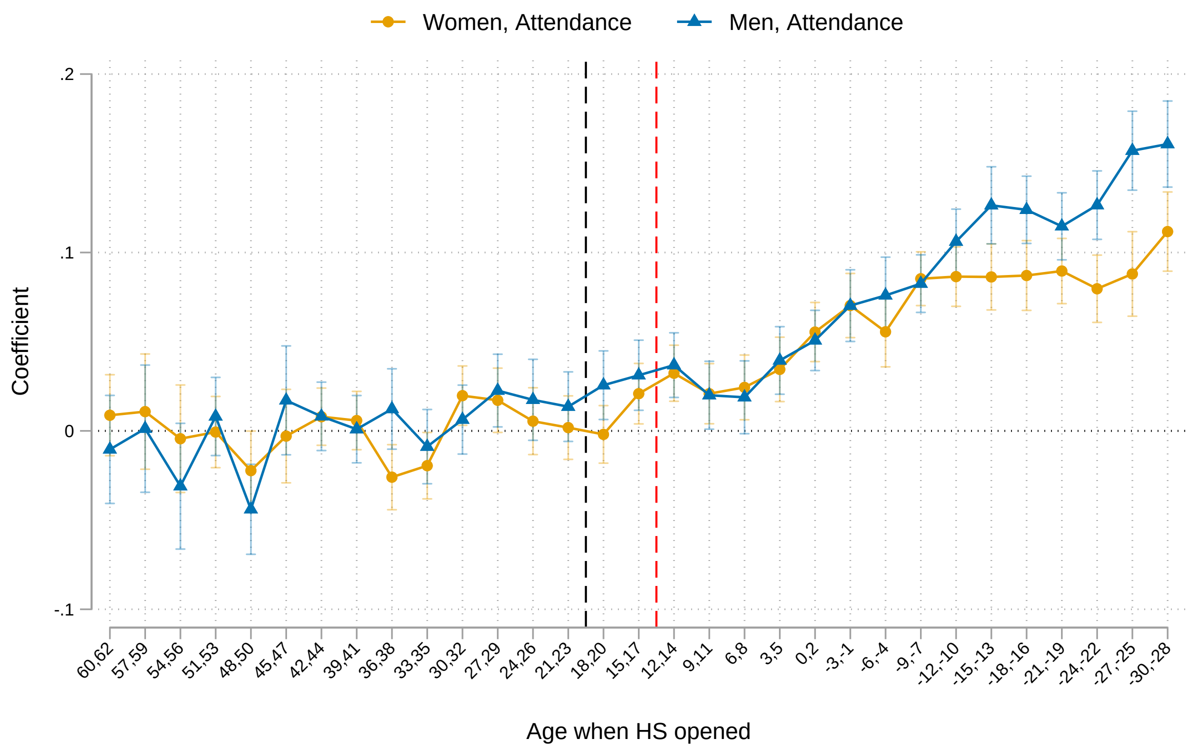
Notes: This map shows the location and founding year of the first high school in U.S. cities and towns between 1800 and 1950. Each dot represents a city, colored by the decade its first high school was established. High schools emerged earliest in New England and the Midwest (lighter colors), spreading later to the South and West. Data compiled from Bureau of Education censuses (1873–1912), accreditation lists (1911–1944), Patterson's Directories (1906–1924), and Private School Universe Surveys (1989–1995).

Figure 2: Effect of high school entry on school attendance, by age



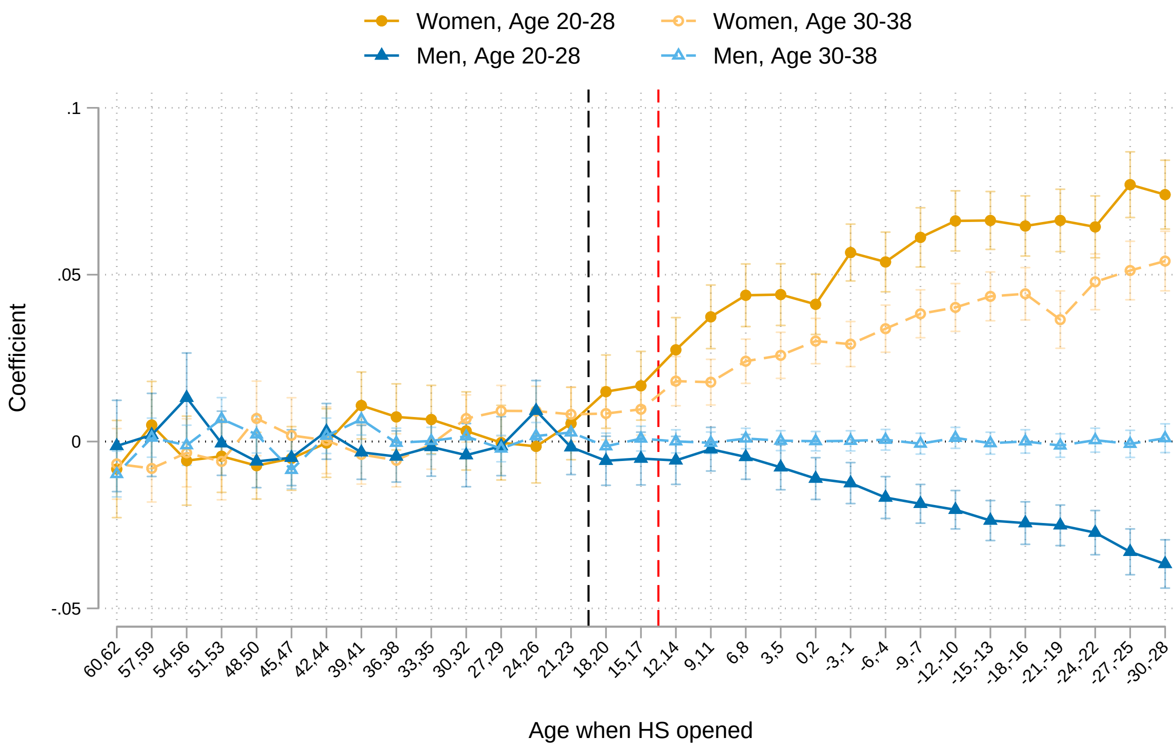
Notes: This figure shows event-study estimates from Equation 1 of the effect of high school access on the probability of reporting school attendance for the indicated age groups. The x-axis represents age when the nearest high school opened. The black dashed line indicates partially treated cohorts; the red dashed line indicates children who could have attended all four traditional years of high school. Treatment group includes children living ≤0.5 miles from a high school; control group lives ≥3 miles away. The model includes city, birth-year, census year, county-by-birth-year, and high-school-opening-year-by-birth-year fixed effects. We normalize pre-period coefficients to the average of pre-treatment coefficients. Standard errors clustered by nearest high school and we show 95 percent confidence intervals. Sample: 11–18 year-olds in the 1850–1930 censuses living in cities whose nearest high school opened before 1909.

Figure 3: Effect of high school entry on 17–18 year-old school attendance, by gender



Notes: These panels show event-study estimates from Equation 1 separately for young women and young men. See Figure 2 notes for full specification details.

Figure 4: Effect of high school entry on labor force participation, by gender and age

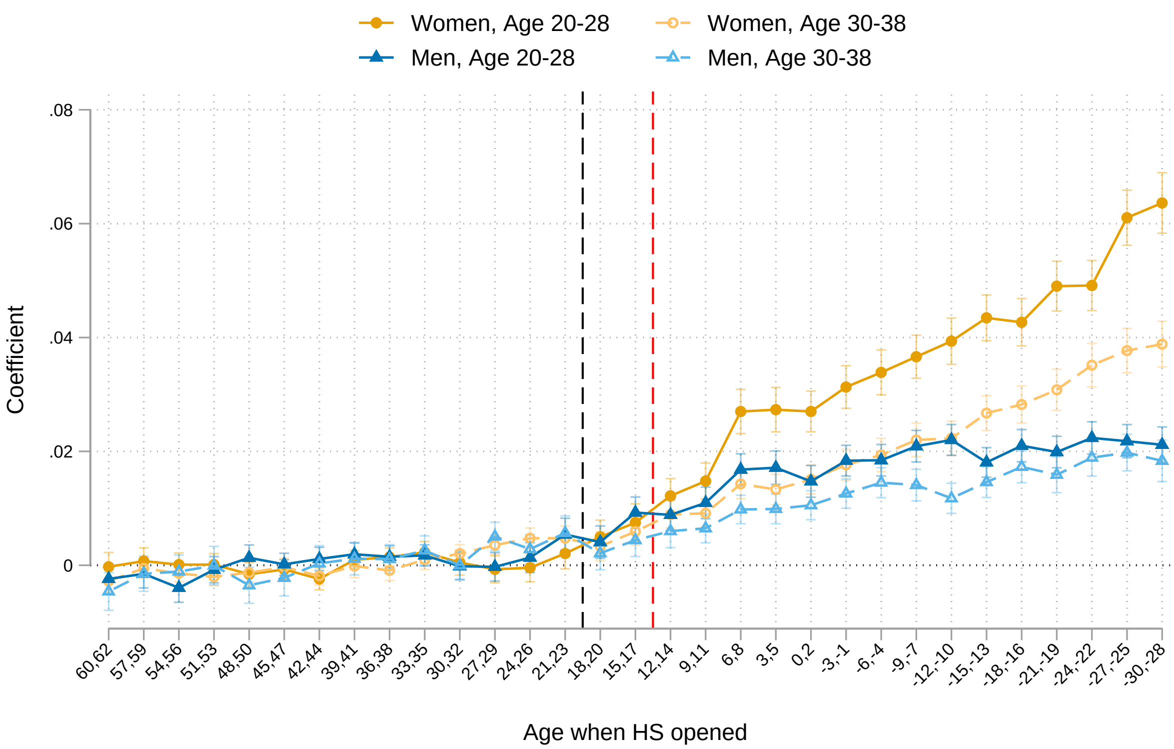


Notes: These panels show event-study estimates from Equation 2 of high school access on labor force participation, measured when linked individuals are ages 20–28 or 30–38. Results shown separately for men and women for each age group. The x-axis represents age when the nearest high school opened. The black dashed line indicates partially treated cohorts; the red dashed line indicates children who could have attended all four traditional years of high school. Treatment group includes children living $\leq$0.5 miles from a high school; control group lives $\geq$3 miles away. The model includes city, birth-year, census year, county-by-birth-year, and high-school-opening-year-by-birth-year fixed effects. Sample: individuals aged 10–18 in childhood cities (1850–1930 censuses) linked to adult outcomes via the Census Tree (Buckles et al. 2025). We normalize pre-period estimates to the average of pre-treatment coefficients. Standard errors clustered by nearest high school and we show 95 percent confidence intervals.

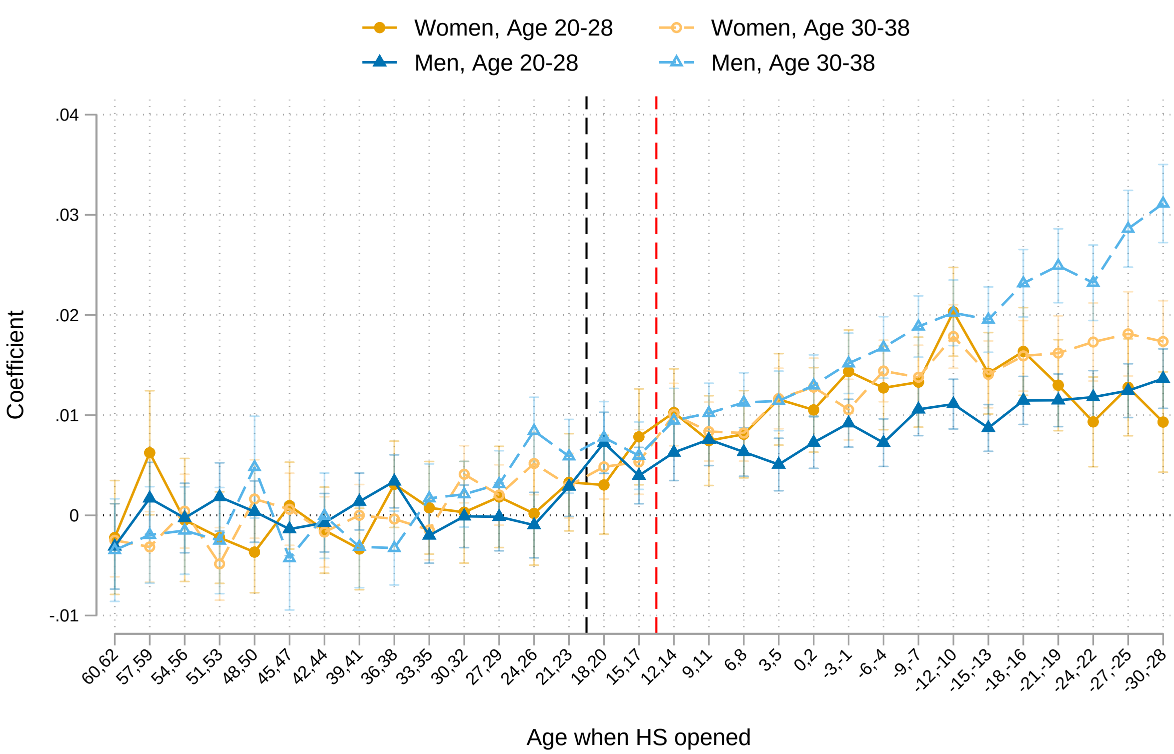
The decline in age 20–28 male labor force participation is consistent with a similar rise in age 20–28 college attendance. See Figure A5.

Figure 5: Effect of high school entry on probability of being in clerical or professional occupations

(a) Clerical occupation



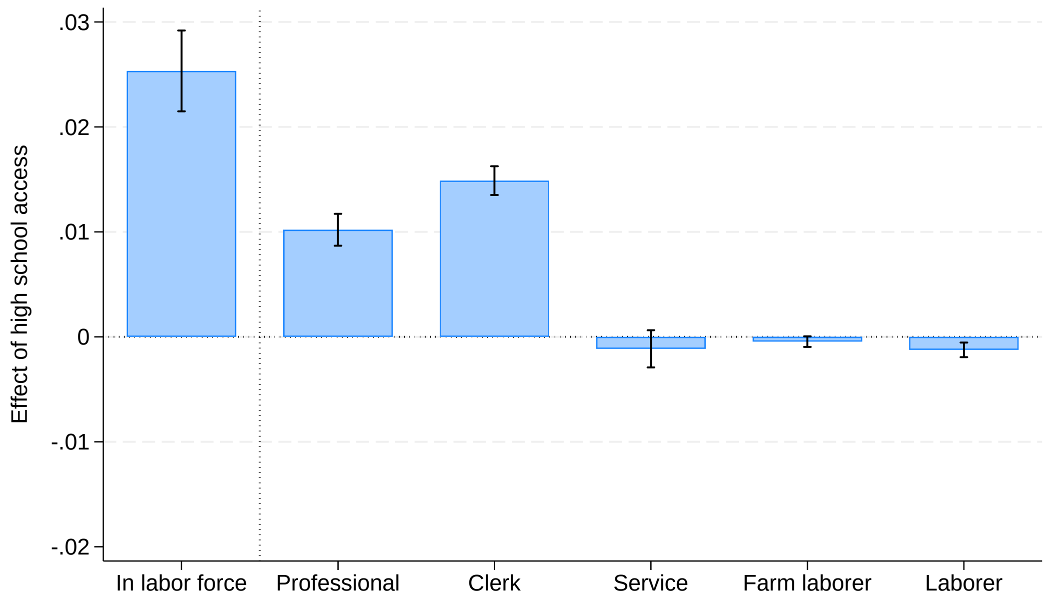
(b) Professional occupation



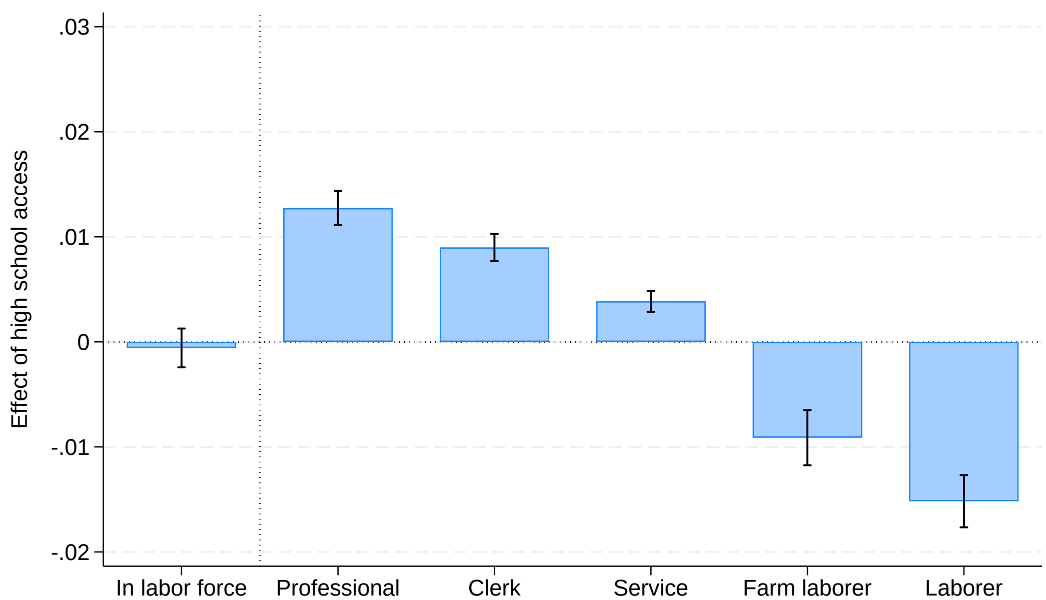
Notes: These panels show event-study estimates from Equation 2 for employment in clerical (Panel A) or professional (Panel B) occupations at age 20-28 or 30-38. See Figure 4 notes for sample and specification details.

Figure 6: Age 30-38 occupation and labor force participation results, by gender

(a) Women



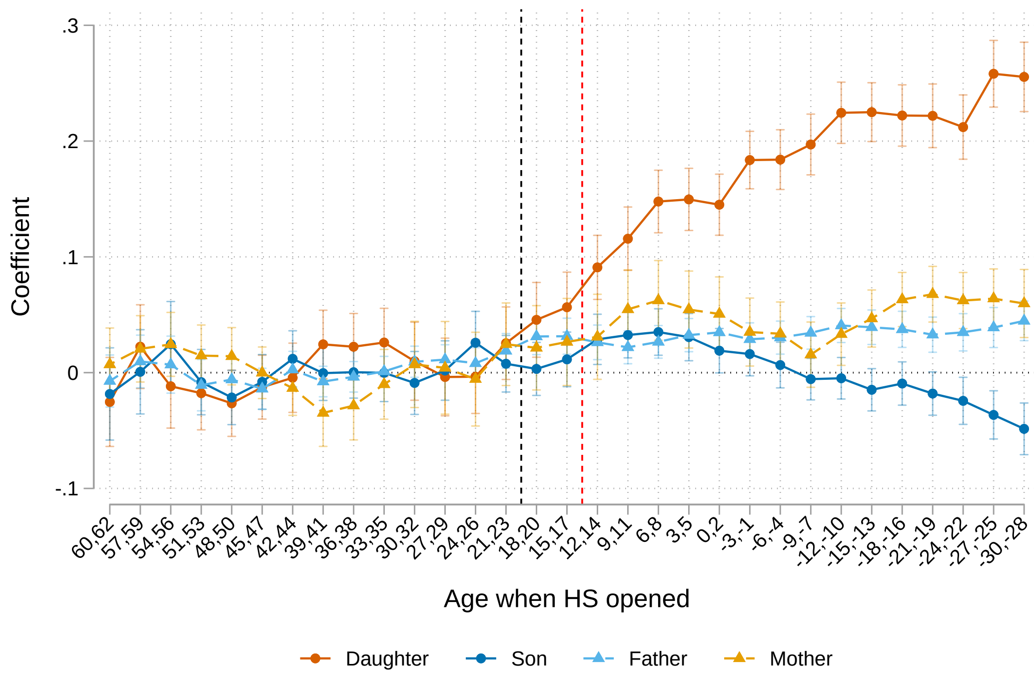
(b) Men



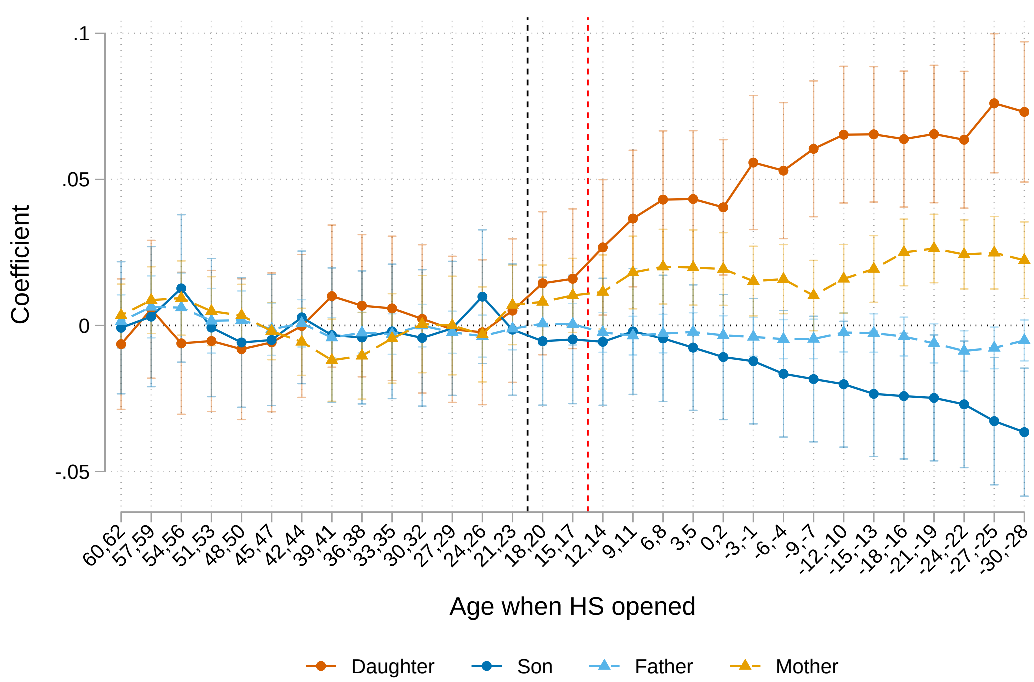
Notes: These figures summarize occupational and labor force effects using a pre-post difference-in-differences specification. Panel A shows results for women; Panel B shows results for men. All coefficients from models with city, birth-year, census year, county-by-birth-year, and high-school-opening-year-by-birth-year fixed effects; see Figure 4 notes for specification and sample details. Standard errors are clustered by nearest high school and 95 percent confidence intervals are shown.

Figure 7: Impact on age 20-28 labor supply and occupational income compared with contemporaneous effects on parent characteristics

(a) Log occupational score



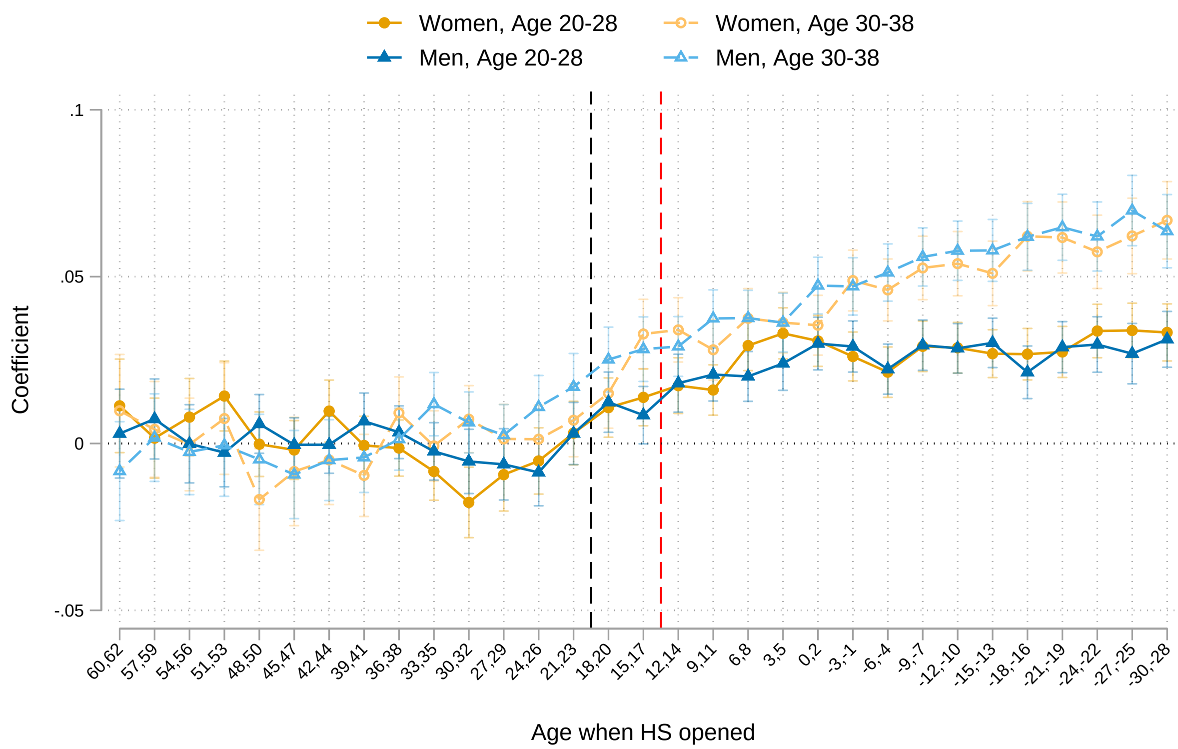
(b) Labor force participation



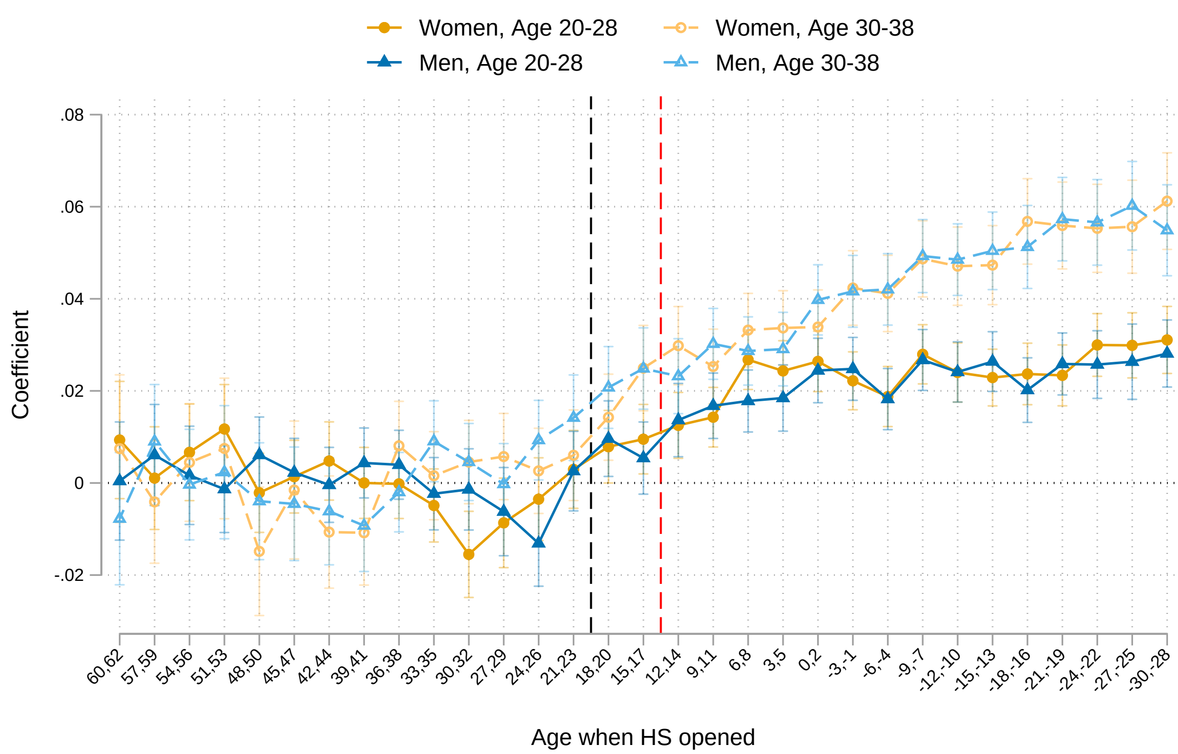
Notes: These panels show event-study estimates from Equation 2 separately for each indicated outcome and sample. See Figure 4 notes for age 20–28 specification and sample details. Results for parents (fathers and mothers) are estimated when youth are age 10-18 in the indicated high school treatment timing cohort.

Figure 8: Effect of high school entry on geographic mobility

(a) Moved 50+ miles from childhood home



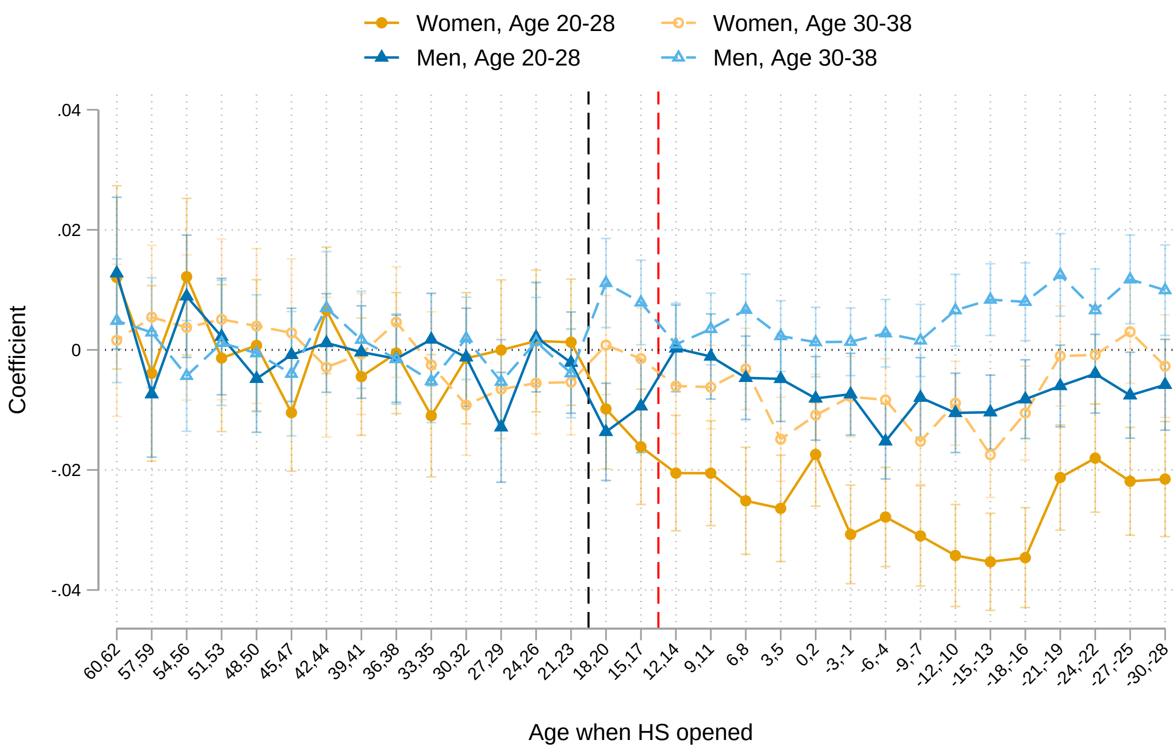
(b) Moved 100+ miles from childhood home



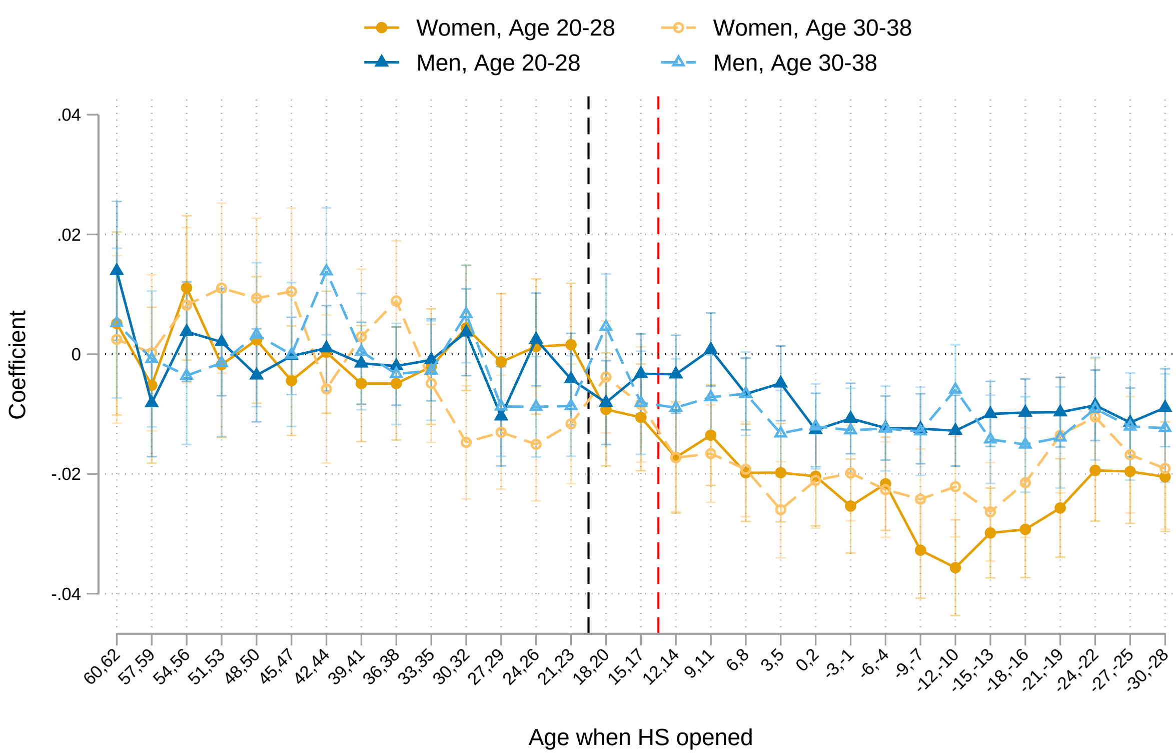
Notes: These panels show event-study estimates from Equation 2 for moving the indicated number of miles from childhood locations measured at age 10-18. See Figure 4 notes for sample and specification details.

Figure 9: Effect of high school entry on marriage and fertility

(a) Marriage

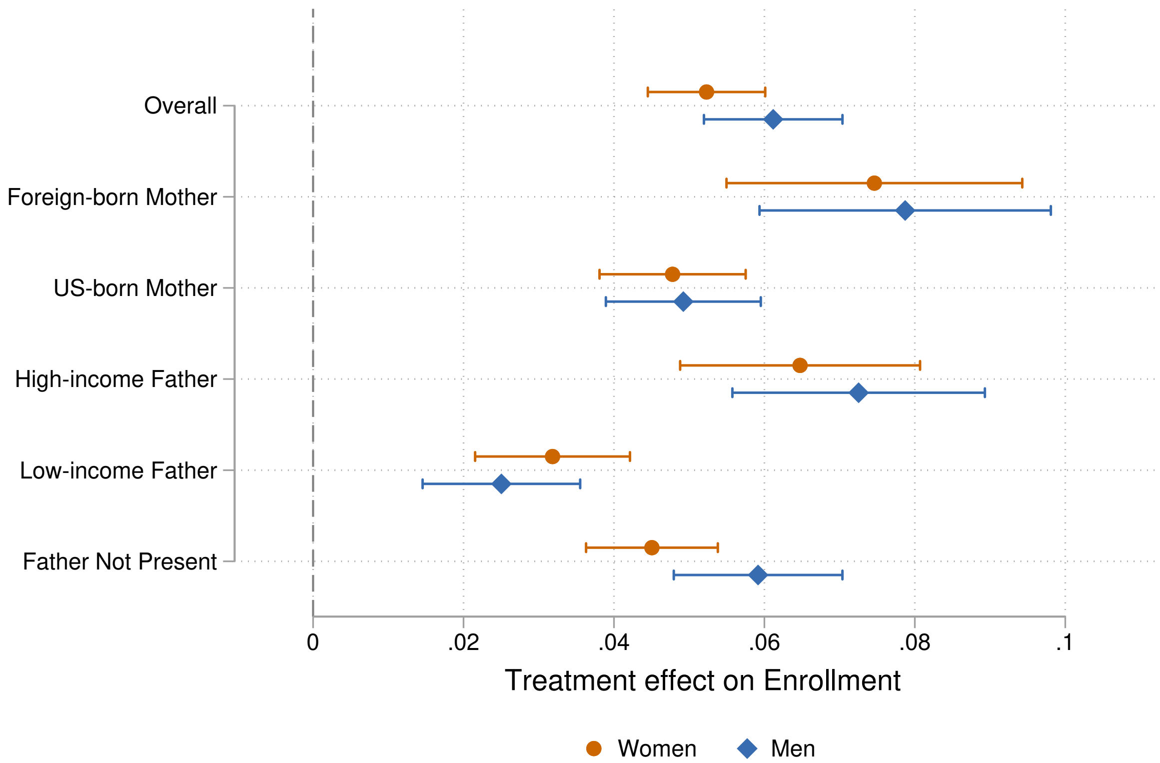


(b) Fertility



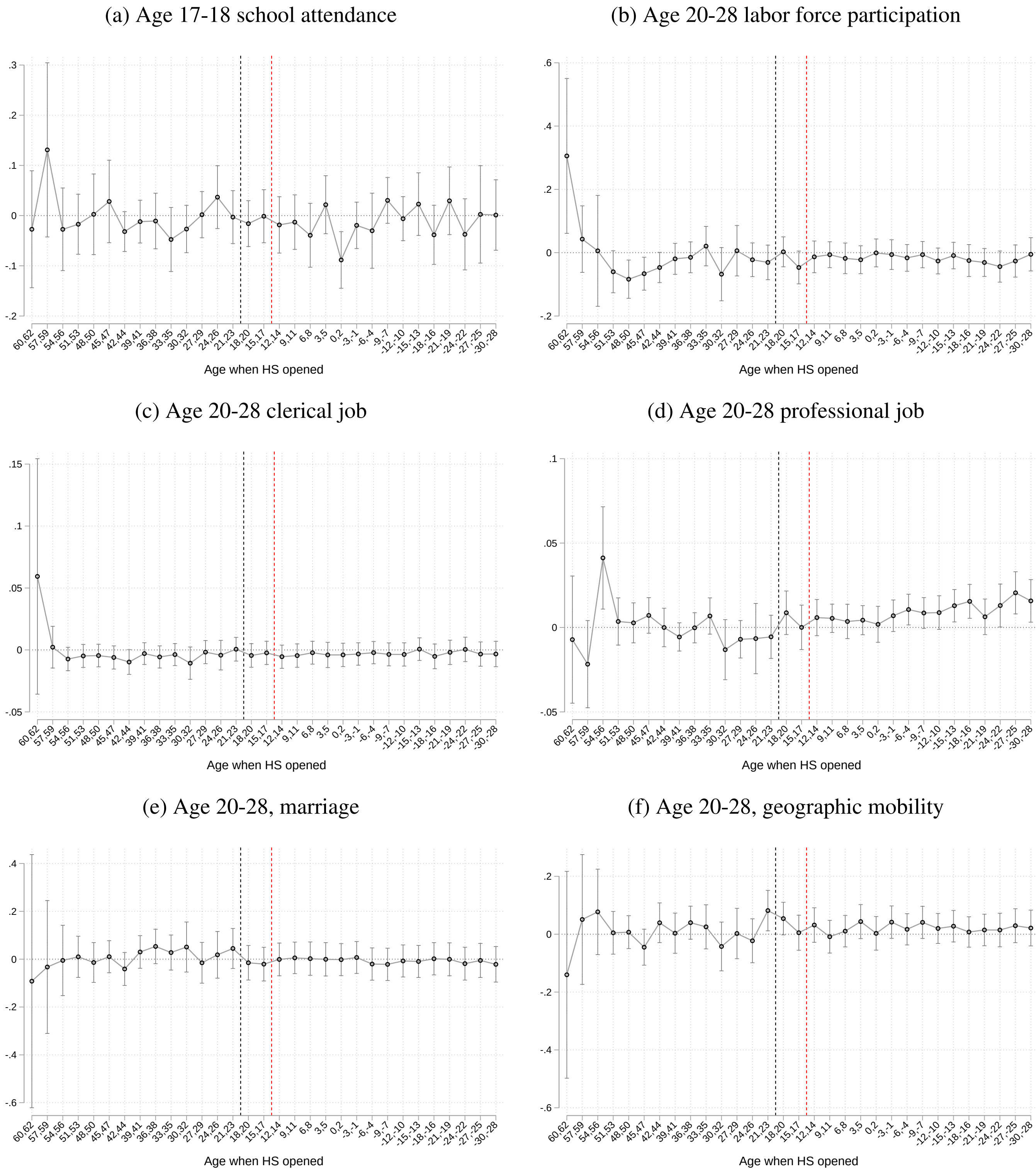
Notes: These panels show event-study estimates from Equation 2 for marriage (Panel A) and fertility, measured by observing the respondent with a child in their household (Panel B). See Figure 4 notes for sample and specification details.

Figure 10: Effect of high school entry on 17-18-year-olds' school attendance, by family background



Notes: This figure shows pre-post difference-in-differences estimates of the effect of high school access on the probability of school attendance for 17–18-year-olds, estimated separately by family background characteristics. The sample is divided by mother's nativity (US- or foreign-born) and father's occupational score (split at the sample median into high- and low-income, with an additional category for children with no father present in the household). Orange circles represent treatment effects for young women, and blue diamonds represent treatment effects for young men. Each coefficient represents a separate regression. Error bars represent 95% confidence intervals. See Figure 2 notes for full specification details.

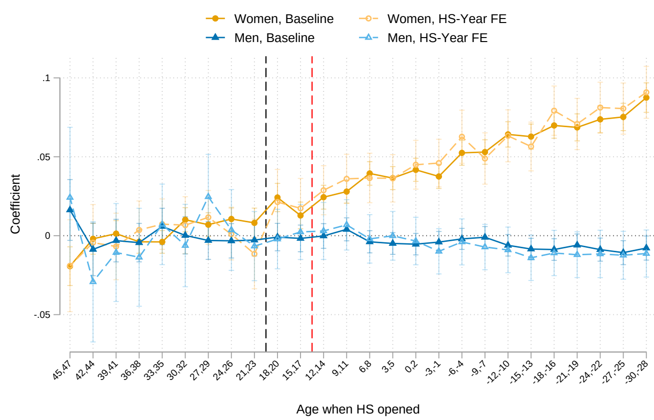
Figure 11: Attendance and labor force results, Black sample



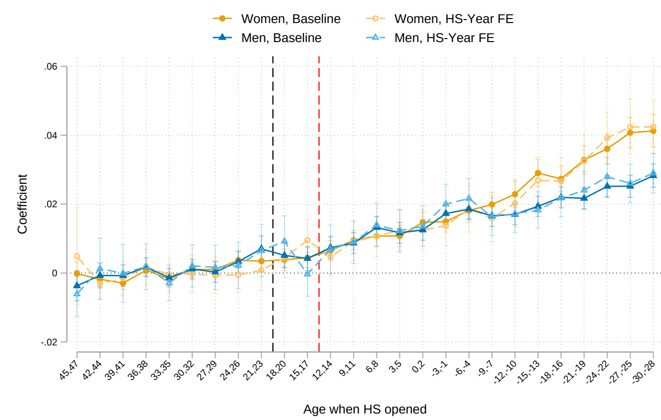
Notes: These panels show event-study estimates from Equations 1 (Panel a) and 2 (Panel B–F) for Black children. Panel A shows attendance effects for 17–18 year-olds. Panels B–F show effects on adult outcomes measured at ages 20–28. See Figures 2 and 4 notes for specification and sample details.

Figure 12: Outcomes measured in 1940, by gender and specification

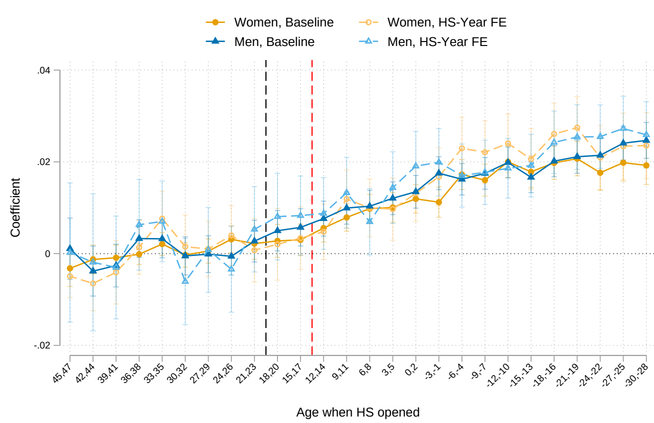
(a) Labor force participation



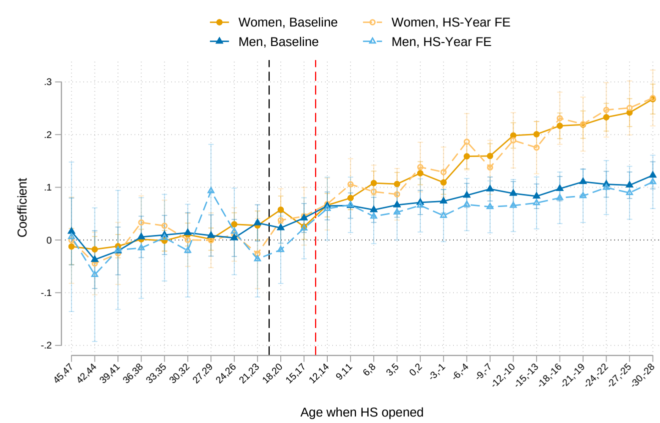
(b) Clerical occupation



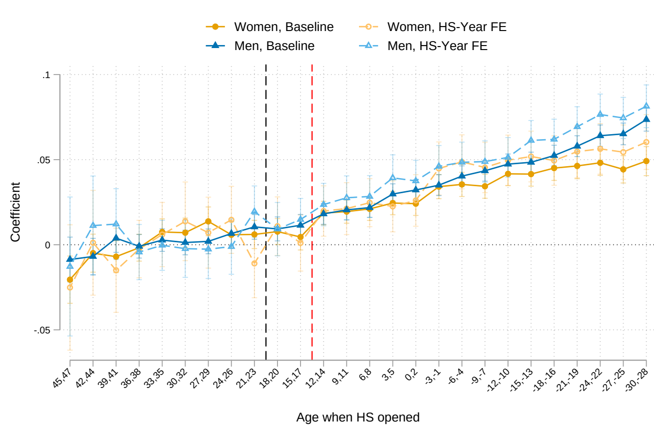
(c) Professional occupation



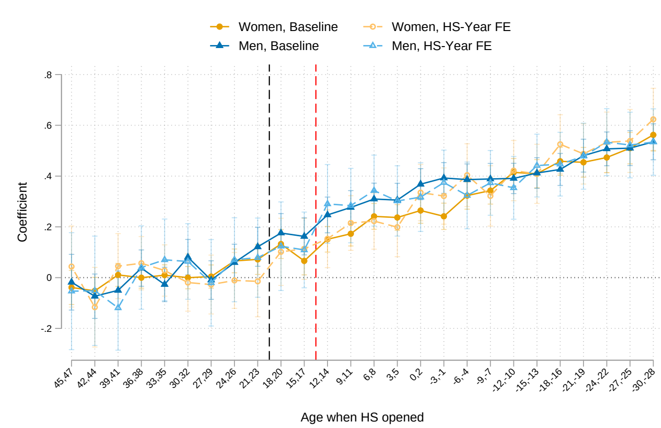
(d) Log occupation score



(e) Attended college



(f) Log reported wage income



Notes: These panels show event-study estimates from Equation 2 for outcomes measured in the 1940 Census for the subsample of individuals who can be linked to 1940. See Figure 4 for full specification and sample details. In addition to the baseline fixed effects, lighter orange and blue lines are estimates from models that additionally condition on specific high school indicator by birth-year fixed effects.

References

Aaronson, Daniel, and Bhashkar Mazumder. 2011. "The Impact of Rosenwald Schools on Black Achievement." *Journal of Political Economy* 119 (5): 821–888.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, Myera Rashid, and Santiago Pérez. 2022. *Census Linking Project: Version 3 [dataset].* https://doi.org/10.7910/DVN/KO5J44.

Andrews, Michael J. 2023. "How do institutions of higher education affect local invention? Evidence from the establishment of US colleges." *American Economic Journal: Economic Policy* 15 (2): 1–41.

Berkes, Enrico, Ezra Karger, and Peter Nencka. 2023. "The Census Place Project: A method for geolocating unstructured place names." *Explorations in Economic History* 87:101477. https://doi.org/10.1016/j.eeh.2022.101477.

Bleemer, Zachary, and Sarah Quincy. 2025. "Junior Colleges and Local Development." Working paper. https://vanderbilt.box.com/s/7ywd2v3rl6cmky0snf1j891xm6q704sw.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley EB Wilbert. 2025. "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project." *Explorations in Economic History,* 101717.

Carruthers, Celeste K., and Marianne H. Wanamaker. 2017. "Separate and unequal in the labor market: Human capital and the Jim Crow wage gap." *Journal of Labor Economics* 35 (3): 655–696.

Cascio, Elizabeth U., and Ethan G. Lewis. 2024. "Teacher Salaries and Racial Inequality in Educational Attainment in the Midcentury South." *Journal of Labor Economics* 42 (S1): S95–S131.

Chancel, Lucas, and Thomas Piketty. 2021. "Global Income Inequality, 1820–2020: the Persistence and Mutation of Extreme Inequality." *Journal of the European Economic Association* 19, no. 6 (October): 3025–3062. ISSN: 1542-4766. https://doi.org/10.1093/jeea/jvab047.

De Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110 (9): 2964–2996.

Derenoncourt, Ellora, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick. 2024. "Wealth of two nations: The US racial wealth gap, 1860–2020." *The Quarterly Journal of Economics* 139 (2): 693–750.

Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91 (4): 795–813.

———. 2004. "The Medium Run Effects of Educational Expansion: Evidence from a Large School Construction Program in Indonesia." *Journal of Development Economics* 74 (1): 163–197.

Goldin, Claudia. 1988. *Marriage Bars: Discrimination Against Married Women Workers, 1920's to 1950's.* Working Paper 2747. National Bureau of Economic Research, October. https://doi.org/10.3386/w2747.

———. 1998. "America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century." *The Journal of Economic History* 58 (2): 345–374.

———. 1999. "Egalitarianism and the Returns to Education during the Great Transformation of American Education." *Journal of Political Economy* 107 (S6): S65–S94.

———. 2006. "The Quiet Revolution that Transformed Women's Employment, Education, and Family." *American Economic Review* 96 (2): 1–21.

Goldin, Claudia, and Lawrence F. Katz. 1998. *Human Capital and Social Capital: The Rise of Secondary Schooling in America, 1910–1940.* Working Paper, Working Paper Series 6439. National Bureau of Economic Research, March. https://doi.org/10.3386/w6439. https://www.nber.org/papers/w6439.

———. 2008. *The Race between Education and Technology.* Harvard University Press.

———. 2011. "Mass Secondary Schooling and the State: The Role of State Compulsion in the High School Movement." In *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy,* edited by Dora L. Costa and Naomi R. Lamoreaux, 275–310. University of Chicago Press / NBER.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics* 225 (2): 254–277.

Helgertz, Jonas, Steven Ruggles, John Robert Warren, Catherine A. Fitch, J. David Hacker, Matt A. Nelson, Joseph P. Price, Evan Roberts, and Matthew Sobek. 2023. *IPUMS Multigenerational Longitudinal Panel: Version 1.1 [dataset].* Minneapolis, MN. https://doi.org/10.18128/D016.V1.1.

Jackson, C. Kirabo, and Claire L. Mackevicius. 2024. "What impacts can we expect from school spending policy? Evidence from evaluations in the United States." *American Economic Journal: Applied Economics* 16 (1): 412–446.

Kaestle, Carl F. 1983. *Pillars of the republic: Common schools and American society, 1780-1860.* Vol. 154. Macmillan.

Li, Sophie. 2025. "Returns to Education for Women in the Mid-Twentieth Century: Evidence from Compulsory Schooling Laws."

Lleras-Muney, Adriana. 2002. "Were Compulsory Attendance and Child Labor Laws Effective? An Analysis from 1915 to 1939." *Journal of Law and Economics* 45 (2): 401–435.

Margo, Robert A. 1986. "Educational Achievement in Segregated School Systems: The Effects of "Separate-but-Equal"." *American Economic Review* 76 (4): 794–801.

Margo, Robert A. 1990. *Race and schooling in the South, 1880-1950: An economic history.* University of Chicago Press.

Navarro-Sola, Laia. 2021. *Secondary Schools with Televised Lessons: The Labor Market Returns of the Mexican Telesecundaria.* Technical report. Working paper.

NCES. 1993. *120 Years of American Education: A Statistical Portrait.* Technical report NCES 93-442. Washington, DC: National Center for Education Statistics, U.S. Department of Education, January.

Parman, John. 2011. "American Mobility and the Expansion of Public Education." *The Journal of Economic History* 71 (1): 105–132.

Reese, William J. 1999. *The origins of the American high school.* Yale University Press.

Ruggles, Steven, Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren. 2024. *IPUMS Ancestry Full Count Data: Version 4.0.* dataset. Minneapolis, MN. https://doi.org/10.18128/D014.V4.0.

Rury, John L. 1991. *Education and Women's Work: Female Schooling and the Division of Labor in Urban America, 1870–1930.* Albany, NY: State University of New York Press.

Schmick, Ethan. 2024. "The Determinants of Early Investments in Urban School Systems in the United States." *Education Finance and Policy* 19 (3): 409–436.

Schmick, Ethan J., and Allison Shertzer. 2019. *The Impact of early investments in urban school systems in the United States.* Technical report. National Bureau of Economic Research.

Smith, James P. 1984. "Race and Human Capital." *American Economic Review* 74 (4): 685–698.

Smith, James P., and Finis Welch. 1986. *Closing the Gap: Forty Years of Economic Progress for Blacks.* Report R-3330-DOL. Santa Monica, CA: Rand Corporation.

———. 1989. "Black Economic Progress After Myrdal." *Journal of Economic Literature* 27 (2): 519–564.

Tyson, Spencer. 2025. "Pathways to Progress: The Role of Academic and Technical High Schools in Black Economic Advancement, 1900-1940." *Available at SSRN 5282550.*

United States Bureau of the Census. 1910. *Thirteenth Census of the United States: Instructions to Enumerators.* Accessed: 2025-12-10. Washington, D.C.: Government Printing Office. https://www.census.gov/history/www/through_the_decades/census_instructions/1910_instructions.html.

United States Department of the Interior. 1870. *Ninth Census, United States. 1870: Instructions to Assistant Marshals.* See instructions for Schedule 1 regarding 'At School' definitions. Washington, D.C.: Government Printing Office. https://www.census.gov/history/www/through_the_decades/census_instructions/1870_instructions.html.

Welch, Finis. 1974. “Education and Racial Discrimination.” In *Discrimination in Labor Markets,* edited by Orley Ashenfelter and Albert Rees, 43–81. Princeton University Press.

A Figures and tables for online publication

Table A1: Summary Statistics

	Full Sample		Women		Men	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Sample Characteristics						
Black	0.100	0.300	0.102	0.303	0.097	0.297
Southern	0.283	0.451	0.286	0.452	0.280	0.449
Age (childhood census)	14.49	2.88	14.52	2.88	14.45	2.88
Distance to HS (miles)	1.41	2.36	1.36	2.33	1.46	2.39
Panel B: Childhood Attendance (Ages 17–18)						
Enrolled in school	0.316	0.465	0.305	0.460	0.327	0.469
Panel C: Long-run Outcomes at Age 20						
In labor force	0.640	0.480	0.296	0.456	0.907	0.290
Clerical occupation	0.050	0.217	0.055	0.229	0.045	0.207
Professional occupation	0.046	0.210	0.061	0.239	0.035	0.183
Log occupational score	1.865	1.484	0.843	1.357	2.657	1.023
Married	0.469	0.499	0.526	0.499	0.425	0.494
Has children	0.300	0.458	0.370	0.483	0.245	0.430
Moved 50+ miles	0.253	0.435	0.224	0.417	0.276	0.447
Panel D: Long-run Outcomes at Age 30						
In labor force	0.639	0.480	0.195	0.396	0.960	0.195
Clerical occupation	0.038	0.191	0.033	0.178	0.042	0.200
Professional occupation	0.044	0.206	0.036	0.186	0.050	0.219
Log occupational score	1.932	1.543	0.569	1.194	2.919	0.869
Married	0.787	0.409	0.777	0.416	0.795	0.404
Has children	0.639	0.480	0.672	0.470	0.614	0.487
Moved 50+ miles	0.369	0.482	0.326	0.469	0.400	0.490
Panel E: Parental Characteristics						
Father in labor force	0.957	0.202	0.956	0.205	0.958	0.200
Father log occ. score	2.810	0.796	2.814	0.806	2.806	0.786
Mother in labor force	0.109	0.312	0.113	0.316	0.106	0.308
Mother log occ. score	0.280	0.848	0.287	0.857	0.272	0.839
Observations	18,381,697		9,182,306		9,199,391	
Places	6,996		6,996		6,996	
High schools	4,846		4,846		4,846	

Notes: Sample includes children observed at ages 10–18 in Census years 1850–1930, living within 0.5 miles (treated) or more than 3 miles (control) of the nearest high school. Long-run outcomes are measured by linking individuals to subsequent Census records. Occupational scores follow IPUMS coding.

Table A2: High School Sample Characteristics

	Mean/Count	SD/Pct
Year of HS opening	1879	18.5
Range	[1780, 1907]	
Distance to HS (treated, miles)	0.058	0.109
Distance to HS (control, miles)	4.95	1.68
HS Opening by Decade		
1780s	4	0.1%
1790s	7	0.1%
1800s	58	1.2%
1810s	50	1.0%
1820s	53	1.1%
1830s	116	2.4%
1840s	161	3.3%
1850s	353	7.3%
1860s	407	8.4%
1870s	803	16.6%
1880s	776	16.0%
1890s	1,320	27.2%
1900s	738	15.2%
Number of high schools	4,846	
Number of places	6,996	

Notes: Sample restricted to high schools opened before 1909 in places observed 6–8 times across Census years. Treated individuals live within 0.5 miles of the nearest high school; control individuals live more than 3 miles away.

Table A3: 1860 City Characteristics by Treatment Status and HS Opening Period

	Pre-1860				1860–1890				1890–1909			
	Treated		Control		Treated		Control		Treated		Control	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A: 1860 City Adult Labor Market Characteristics												
In labor force	0.485	0.081	0.469	0.135	0.487	0.114	0.473	0.138	0.488	0.127	0.477	0.142
Clerical occupation	0.005	0.005	0.002	0.010	0.003	0.005	0.001	0.002	0.002	0.005	0.001	0.002
Professional occupation	0.026	0.013	0.017	0.013	0.025	0.018	0.014	0.014	0.022	0.016	0.012	0.012
Teacher	0.008	0.006	0.009	0.008	0.008	0.012	0.007	0.010	0.008	0.011	0.006	0.007
Service occupation	0.045	0.057	0.043	0.089	0.037	0.061	0.031	0.077	0.041	0.077	0.034	0.083
Farm laborer	0.024	0.027	0.051	0.048	0.030	0.037	0.049	0.053	0.038	0.043	0.045	0.048
Laborer	0.066	0.036	0.046	0.039	0.065	0.053	0.044	0.054	0.053	0.046	0.045	0.053
Log occ. score	2.345	0.191	2.031	0.279	2.268	0.267	2.001	0.339	2.186	0.265	2.002	0.336
Panel B: 1860 City Adult Demographic Characteristics												
Married	0.649	0.069	0.698	0.070	0.661	0.093	0.722	0.081	0.675	0.082	0.715	0.088
Has children	0.559	0.072	0.596	0.072	0.581	0.094	0.640	0.090	0.601	0.083	0.640	0.091
Adult population	4975.9	4084.5	637.5	772.1	1911.3	1725.0	464.3	711.8	1094.2	1053.2	427.4	557.2
Observations	2,555,618		346,783		6,780,010		2,010,717		3,969,285		2,719,284	
High schools	698		293		1,475		1,027		1,197		1,260	

Notes: City-level characteristics measured from 1860 Census for adults aged 19–65. Treated = individuals living <0.5 miles from nearest HS; Control = individuals living >3 miles away. Columns show means by treatment status within each period of high school opening. Pre-1860 includes schools opening before 1860; 1860–1890 includes schools opening 1860–1889; 1890–1909 includes schools opening 1890–1908.

Table A4: Effects of High School Access on Attendance

	Pre-Post		Cohort Comparison		1860	
	Effect	(SE)	Effect	(SE)	Mean	N
Women	0.052	(0.004)	0.027	(0.007)	0.269	1,908,506
Men	0.061	(0.005)	0.007	(0.008)	0.398	1,839,145

Notes: Dependent variable is school attendance for ages 17–18. Pre-Post compares cohorts age 0–20 when HS opened (post) to older cohorts. Cohort Comparison compares ages 0–14 (young) to ages 18–32 (old) when HS opened. Treatment: $\leq$0.5 mi from HS town; Control: $\geq$3 mi. All models include city, birth-year, census year, county-by-birth-year, and HS-cohort-times-birth-year fixed effects. 1860 Mean shows attendance rates from the 1860 Census. Standard errors clustered by nearest high school.

Table A5: Effects of High School Access on Women's Adult Outcomes

	Pre-Post		Cohort Comparison		1860	
	Effect	(SE)	Effect	(SE)	Mean	N
Panel A: Outcomes at Ages 20–28						
Labor Force Participation	0.049	(0.003)	0.023	(0.005)	0.138	3,303,621
Clerical Occupation	0.032	(0.001)	0.014	(0.001)	0.001	3,303,621
Professional Occupation	0.011	(0.001)	0.007	(0.003)	0.033	3,303,621
Log Occupational Score	0.165	(0.008)	0.086	(0.014)	0.381	3,303,621
Married	-0.024	(0.003)	-0.020	(0.005)	0.495	3,303,621
Moved 50+ Miles	0.026	(0.002)	0.017	(0.004)	0.240	3,247,664
Panel B: Outcomes at Ages 30–38						
Labor Force Participation	0.025	(0.002)	0.018	(0.003)	0.182	2,716,781
Clerical Occupation	0.015	(0.001)	0.010	(0.001)	0.001	2,716,781
Professional Occupation	0.010	(0.001)	0.006	(0.001)	0.017	2,716,781
Log Occupational Score	0.085	(0.006)	0.052	(0.008)	0.577	2,716,781
Married	-0.005	(0.002)	-0.004	(0.003)	0.775	2,716,781
Moved 50+ Miles	0.039	(0.003)	0.030	(0.004)	0.320	2,675,662

Notes: Sample includes women observed at ages 10–18 in childhood, linked to adult outcomes. Pre-Post compares cohorts age 0–20 when HS opened to older cohorts. Cohort Comparison compares ages 0–14 to ages 18–32 when HS opened. Treatment: $\leq$0.5 mi from HS; Control: $\geq$3 mi. All models include city, birth-year, census year, county-by-birth-year, and HS-cohort-times-birth-year FEs. 1860 Mean shows outcomes for children observed in the 1860 Census, linked to their adult records. SE clustered by HS.

Table A6: Effects of High School Access on Men's Adult Outcomes

	Pre-Post		Cohort Comparison		1860	
	Effect	(SE)	Effect	(SE)	Mean	N
Panel A: Outcomes at Ages 20–28						
Labor Force Participation	-0.015	(0.002)	-0.008	(0.004)	0.862	4,279,627
Clerical Occupation	0.016	(0.001)	0.011	(0.001)	0.018	4,279,627
Professional Occupation	0.008	(0.001)	0.005	(0.002)	0.024	4,279,627
Log Occupational Score	0.002	(0.006)	0.028	(0.013)	2.466	4,279,627
Married	-0.007	(0.002)	0.003	(0.004)	0.373	4,279,627
Moved 50+ Miles	0.024	(0.003)	0.017	(0.005)	0.321	4,188,887
Panel B: Outcomes at Ages 30–38						
Labor Force Participation	-0.001	(0.001)	0.000	(0.001)	0.962	3,775,119
Clerical Occupation	0.009	(0.001)	0.007	(0.001)	0.021	3,775,119
Professional Occupation	0.013	(0.001)	0.006	(0.001)	0.042	3,775,119
Log Occupational Score	0.035	(0.004)	0.031	(0.006)	2.899	3,775,119
Married	0.006	(0.002)	0.004	(0.003)	0.805	3,775,119
Moved 50+ Miles	0.040	(0.003)	0.031	(0.004)	0.420	3,706,561

Notes: Sample includes men observed at ages 10–18 in childhood, linked to adult outcomes. See Table A5 for specification details. 1860 Mean shows outcomes for children observed in the 1860 Census, linked to their adult records. SE clustered by HS.

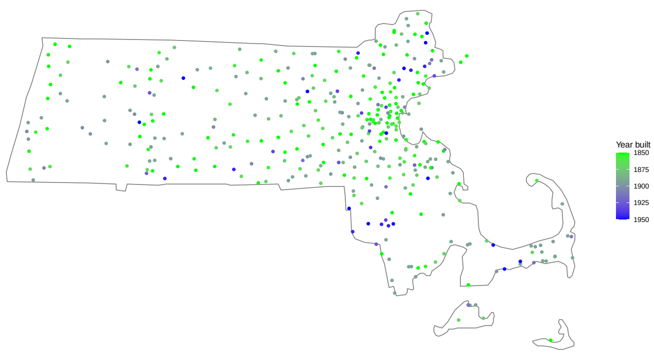
Table A7: Effects of High School Access on Parent Characteristics

	Pre-Post		Cohort Comparison		1860	
	Effect	(SE)	Effect	(SE)	Mean	N
Panel A: Father Characteristics						
Father in Labor Force	-0.002	(0.001)	-0.003	(0.002)	0.886	12869877
Father Log Occ. Score	0.029	(0.004)	0.012	(0.007)	2.604	12869877
Panel B: Mother Characteristics						
Mother in Labor Force	0.019	(0.004)	0.009	(0.007)	0.083	12869877
Mother Log Occ. Score	0.046	(0.011)	0.022	(0.024)	0.200	12869877

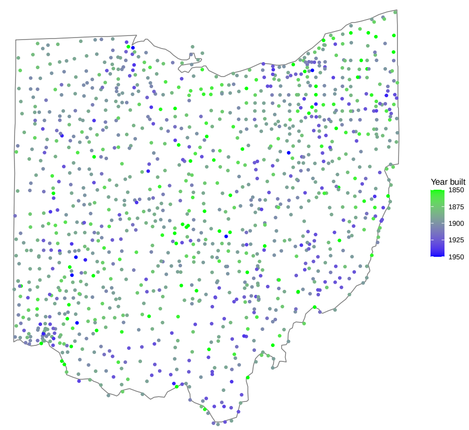
Notes: Tests whether HS openings predict changes in contemporaneous parent characteristics. Sample: children ages 10–18, pooled across gender. All models include city, birth-year, census year, county-by-birth-year, and HS-cohort-by-birth-year FEs. 1860 Mean from the 1860 Census. SE clustered by HS.

Figure A1: Map of cities, by year of first high school constructed (selected states)

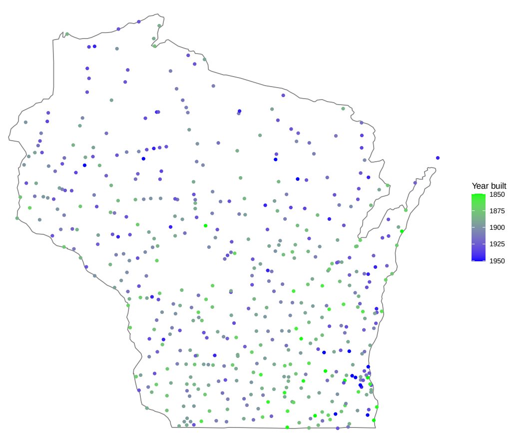
(a) Massachusetts



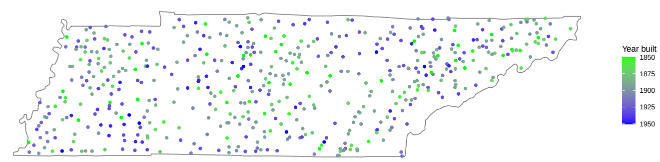
(b) Ohio



(c) Wisconsin



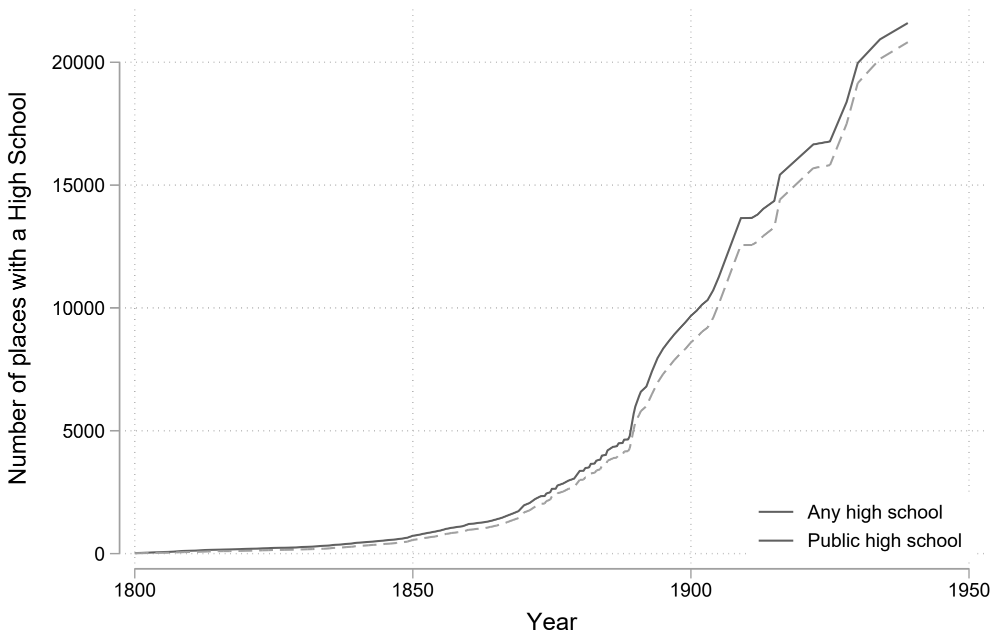
(d) Tennessee



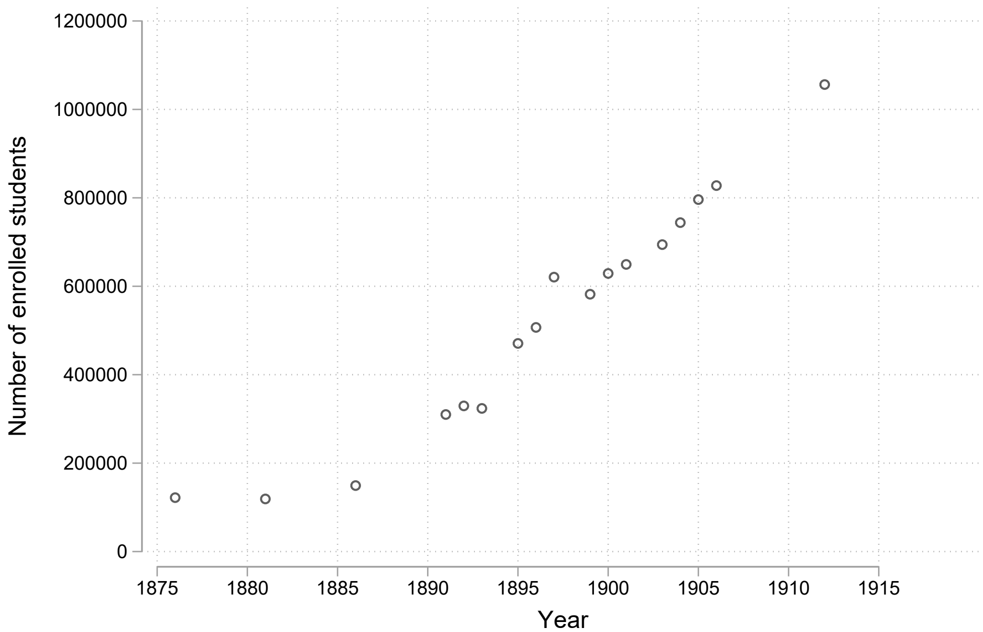
Notes: This map shows the founding dates of the first high school built in each town we can geolocate in Massachusetts (a), Ohio (b), Wisconsin (c), and Tennessee (d).

Figure A2: Towns with high schools and overall attendance

(a) Number of towns in the U.S. with high schools, by year



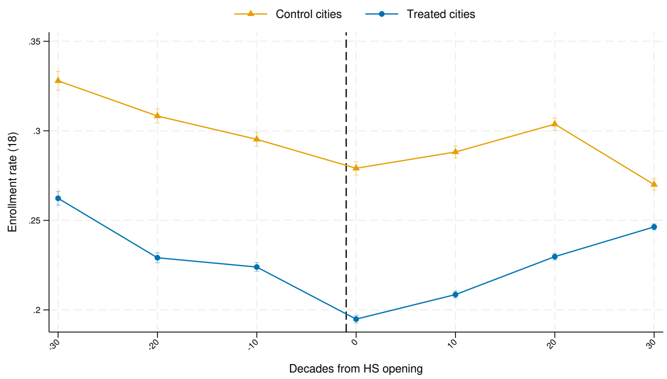
(b) Number of students enrolled in high schools



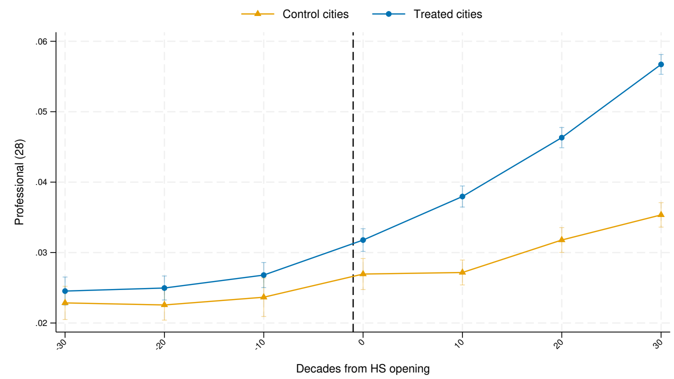
Notes: Panel A shows the cumulative number of U.S. cities and towns with at least one high school from 1800 to 1950. The solid line includes both public and private high schools; the dashed line includes only public high schools. Flat segments after 1910 reflect gaps in data availability rather than stagnant growth. Our main analysis focuses on high schools established before 1909, when annual data quality is highest. Panel B shows total high school attendance for years with available data.

Figure A3: Average trends for youth in treatment and control cities

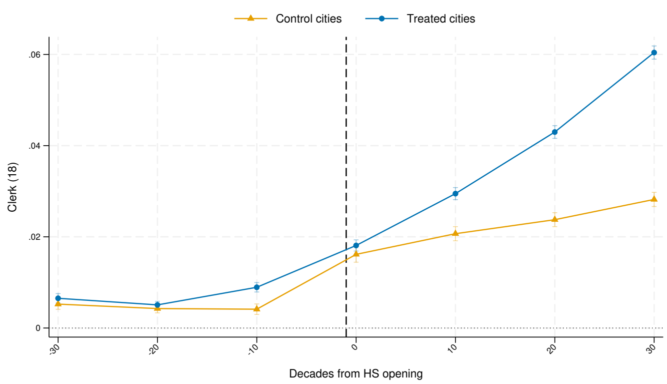
(a) School attendance (age 18)



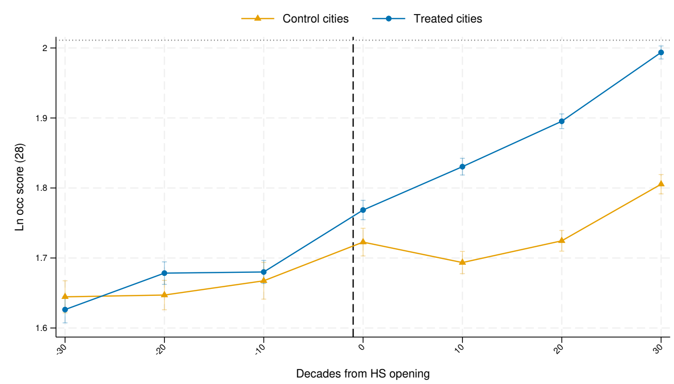
(b) Professional worker (age 28)



(c) Clerical worker (age 28)



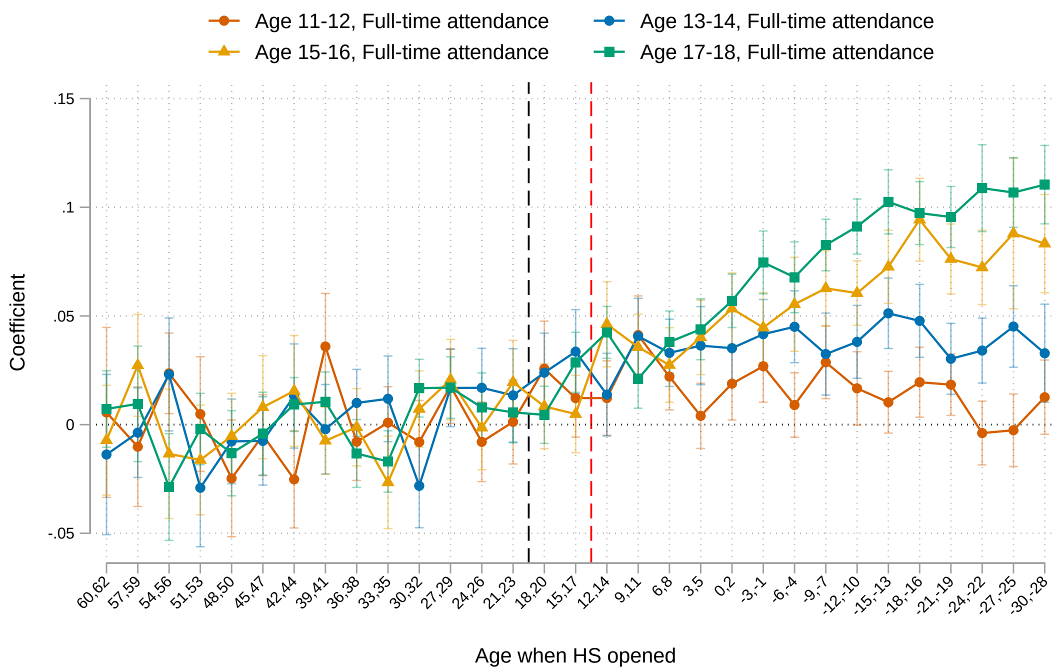
(d) Log occupation score (age 28)



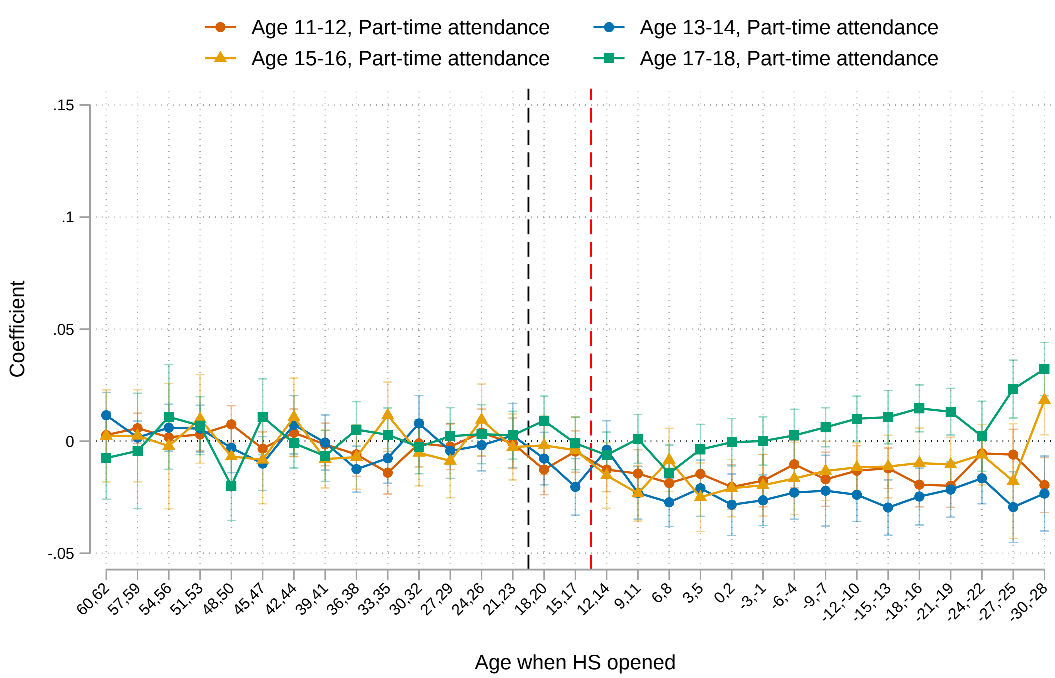
Notes: These panels show the average and standard-deviation derived 95-percent confidence interval for each indicated outcome. Treatment group includes children living $\leq$0.5 miles from a high school; control group lives $\geq$3 miles away. Averages are shown in the decennial censuses leading up to when their nearest high school was built; treatment timing in this setting is well defined for both treatment and control cities. Enrollment is measured at age 18. Age 28 outcomes are measured by linking forward youth 10 years using the Census Tree (Buckles et al. 2025).

Figure A4: Effect of high school entry on ‘full’ and ‘part-time’ school attendance, by age

(a) ‘Full-time’ attendance

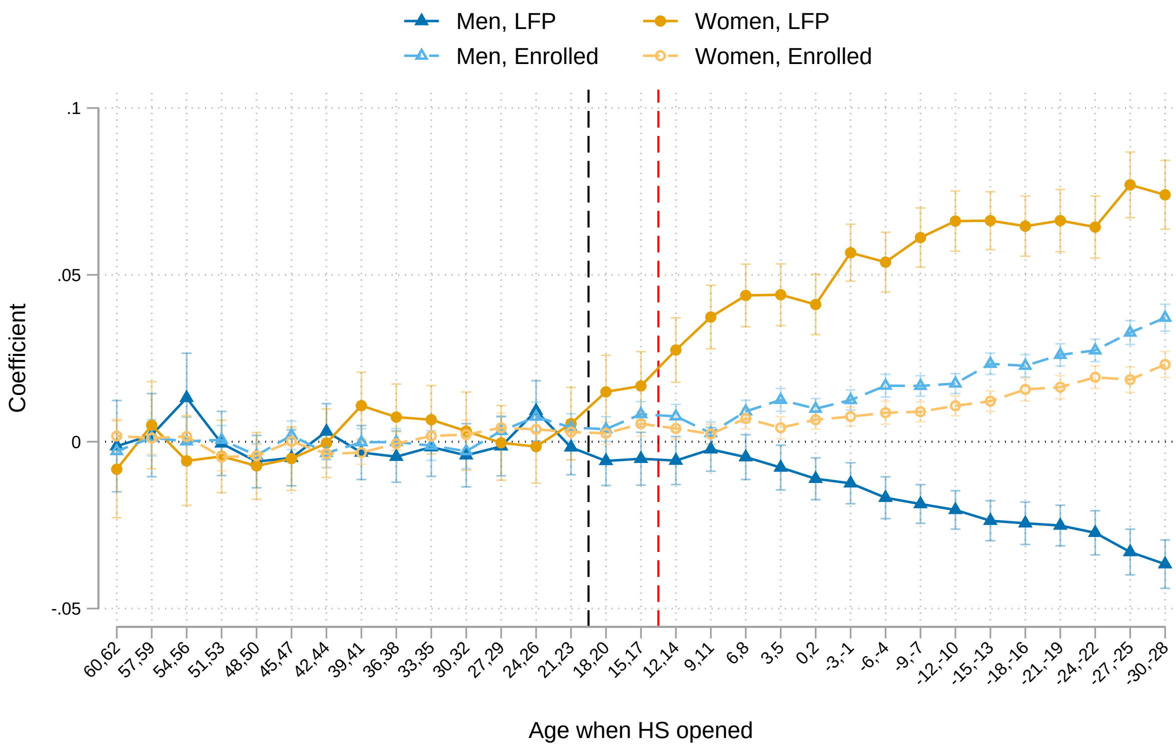


(b) ‘Part-time’ attendance



Notes: This figure shows event-study estimates from Equation 1 of the effect of high school access on the probability of reporting school attendance for the indicated age groups. ‘Full-time’ attendance is an indicator equal to 1 if a respondent reported attending school and not having an occupation. ‘Part-time’ attendance is an indicator equal to 1 if a respondent reported attending school and reported having an occupation. The x-axis represents age when the nearest high school opened. The black dashed line indicates partially treated cohorts; the red dashed line indicates children who could have attended all four traditional years of high school. Treatment group includes children living $\leq$0.5 miles from a high school; control group lives $\geq$3 miles away. The model includes city, birth-year, census year, county-by-birth-year, and high-school-opening-year-by-birth-year fixed effects. Standard errors clustered by nearest high school. Sample: 11–18 year-olds in the 1850–1930 censuses living in cities whose nearest high school opened before 1909.

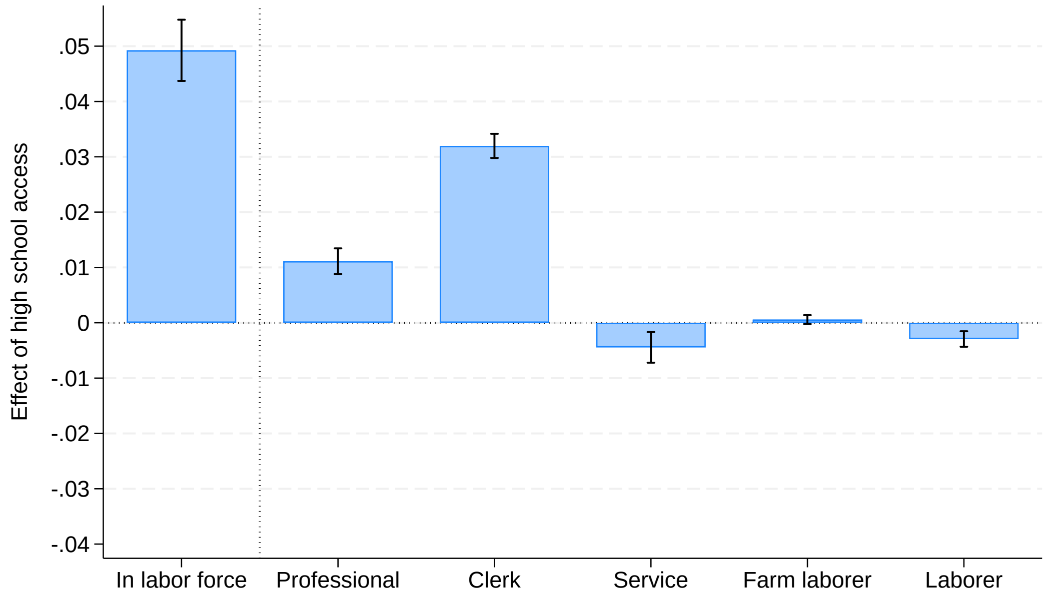
Figure A5: High school impact on age 20s labor force participation compared to impact on age 20s school attendance



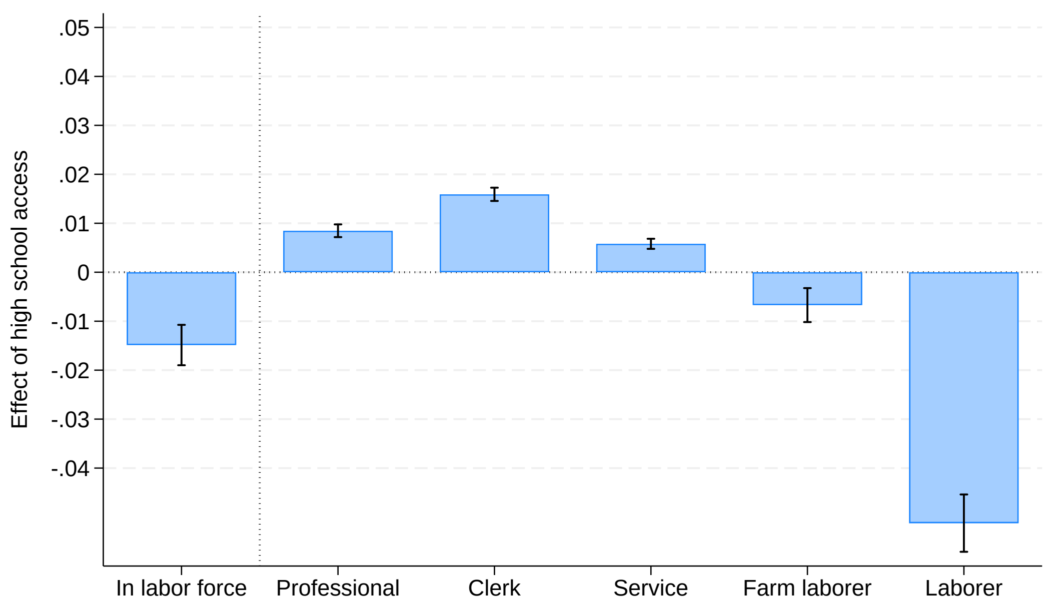
Notes: These panels show event-study estimates from Equation 2 of high school access on labor force participation and reported school attendance, measured when linked individuals are ages 20–28. Results shown separately for men and women. The x-axis represents age when the nearest high school opened. The black dashed line indicates partially treated cohorts; the red dashed line indicates children who could have attended all four traditional years of high school. Treatment group includes children living $\leq$0.5 miles from a high school; control group lives $\geq$3 miles away. The model includes city, birth-year, census year, county-by-birth-year, and high-school-opening-year-by-birth-year fixed effects. Sample: individuals aged 10–18 in childhood cities (1850–1930 censuses) linked to adult outcomes via Census Tree (Buckles et al. 2025). Standard errors clustered by nearest high school.

Figure A6: Age 20-28 occupation and labor force participation results, by gender

(a) Women



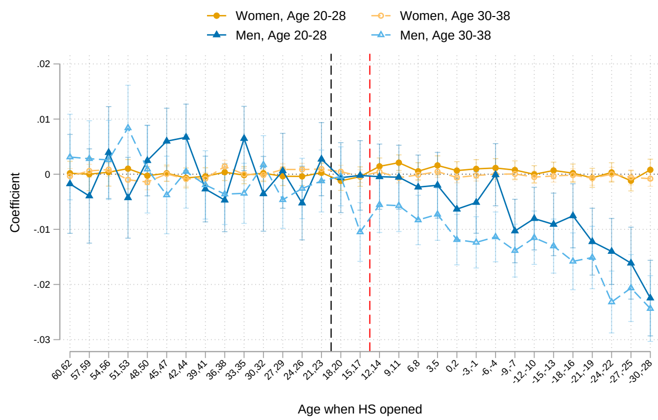
(b) Men



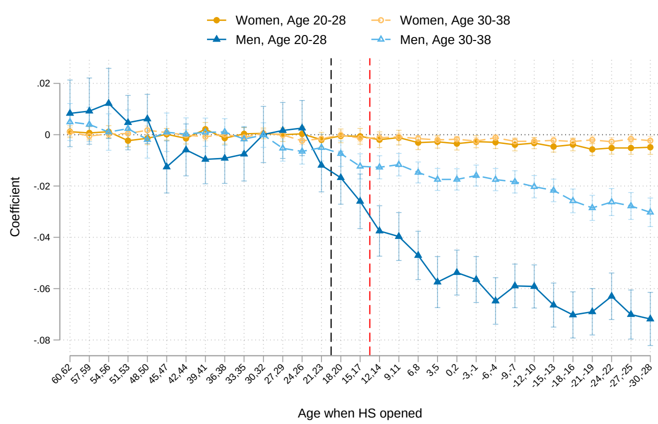
Notes: These figures show pre-post difference-in-differences estimates for occupational and labor force outcomes measured at ages 20–28. Panel A shows results for women; Panel B shows results for men. See Figure 6 notes for specification details.

Figure A7: Effect of high school entry on probability in laborer, farm laborer, and farm owner occupations, by gender and age

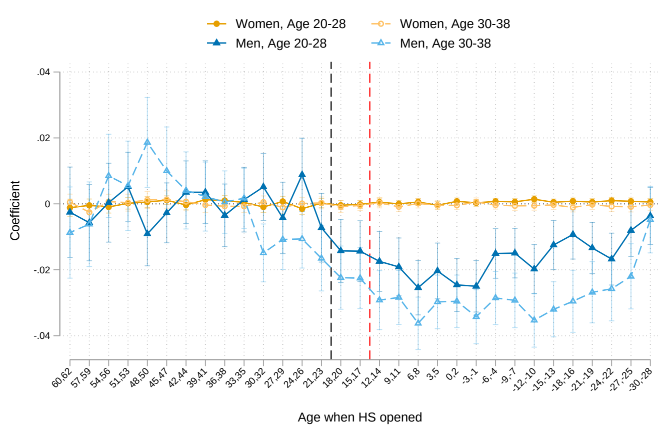
(a) Laborer occupation



(b) Farm laborer occupation



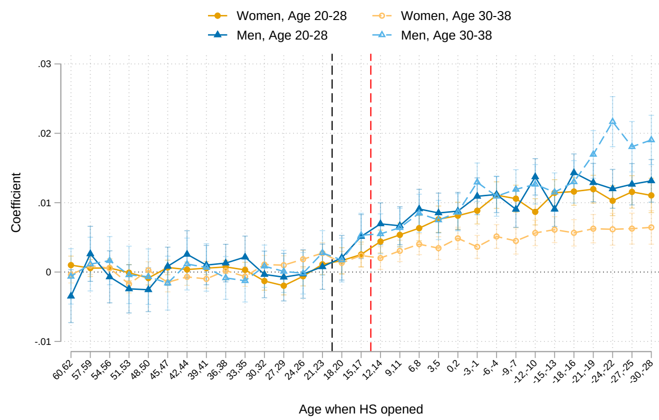
(c) Farm owner



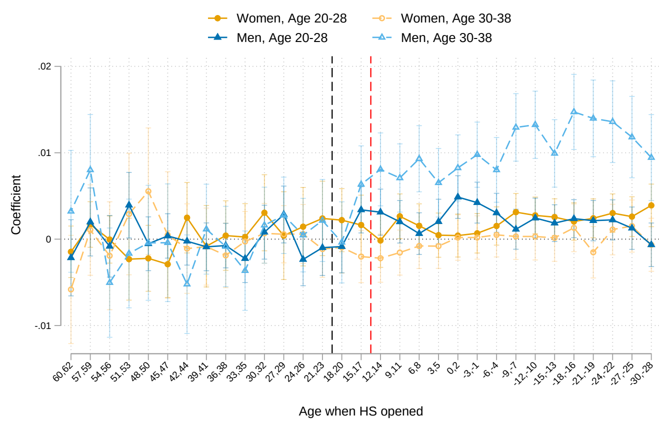
Notes: These panels show event-study estimates from Equation 2 for employment in the indicated occupations. See Figure 4 notes for sample and specification details.

Figure A8: Effect of high school entry on probability in sales, managerial, and service occupations, by gender and age

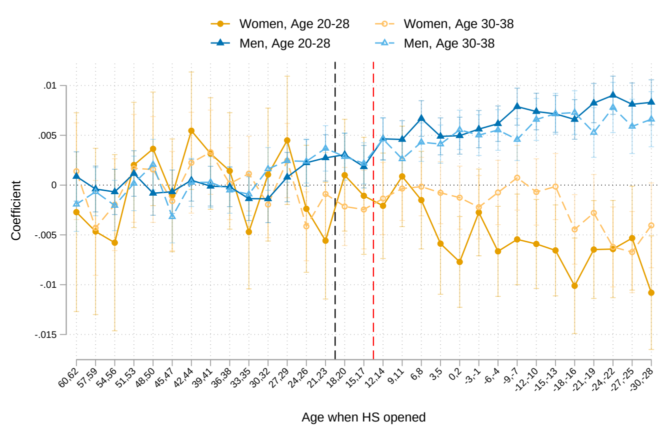
(a) Sales occupation



(b) Manager occupation



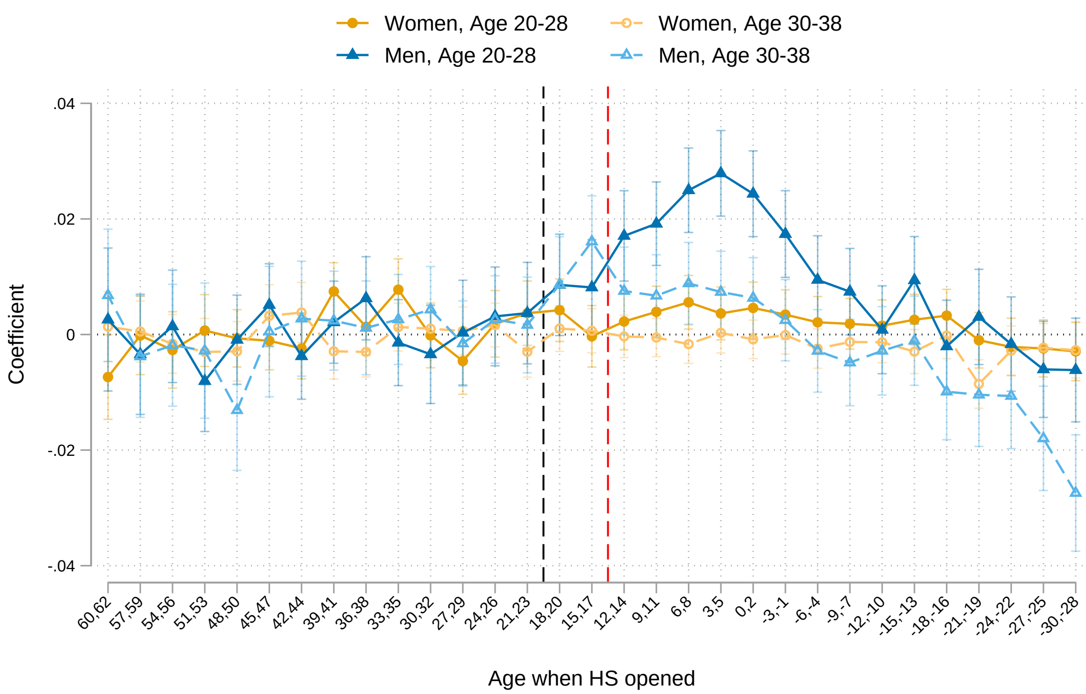
(c) Service occupation



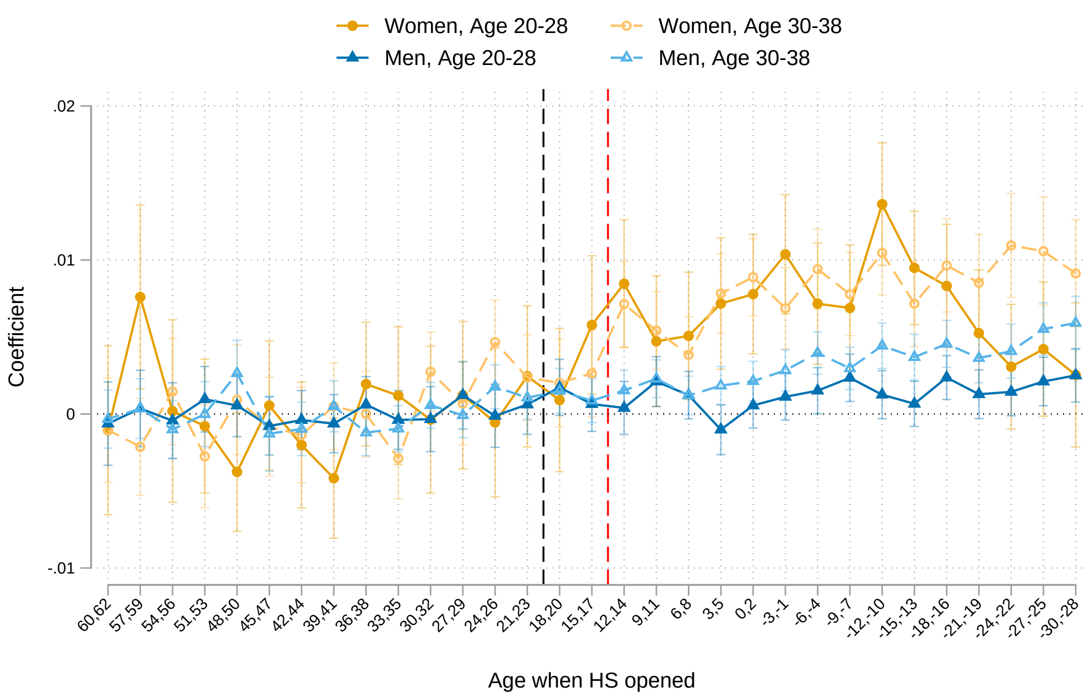
Notes: These panels show event-study estimates from Equation 2 for employment in the indicated occupations. See Figure 4 notes for sample and specification details.

Figure A9: Effect of high school entry on probability in craftsmen/operative and teaching occupations, by gender and age

(a) Craftsmen/operative occupation



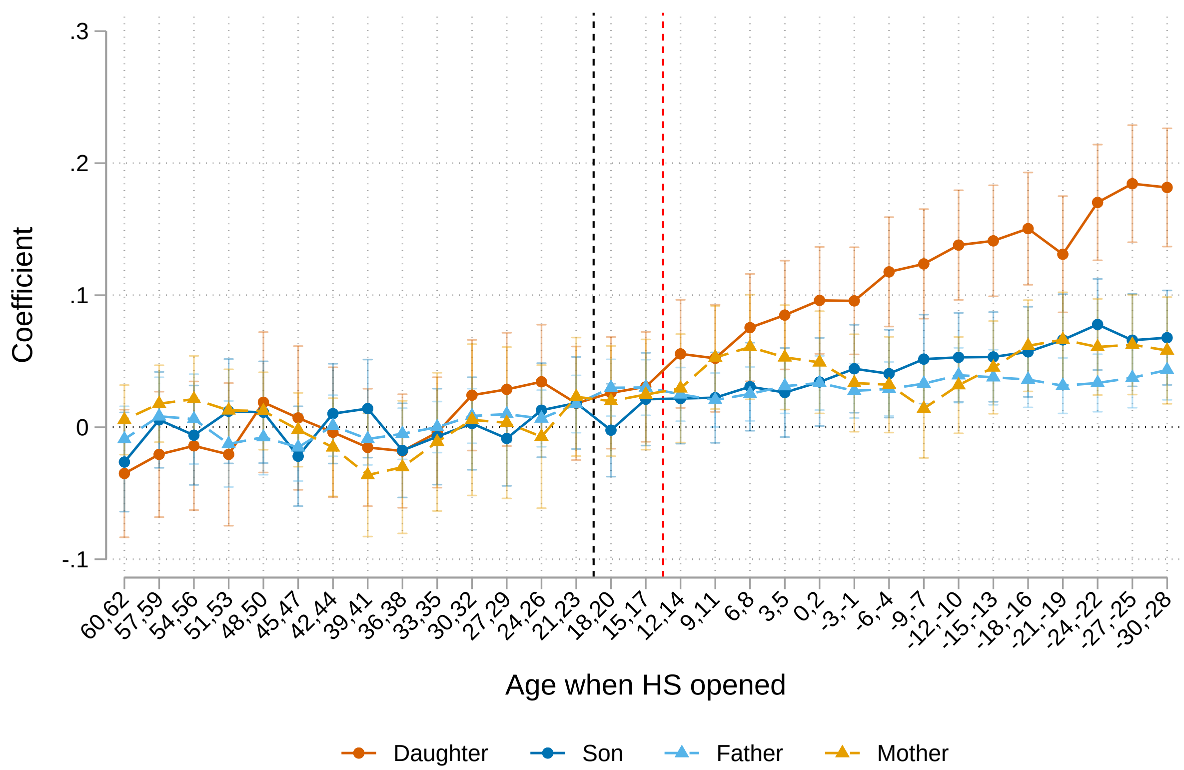
(b) Teaching occupation



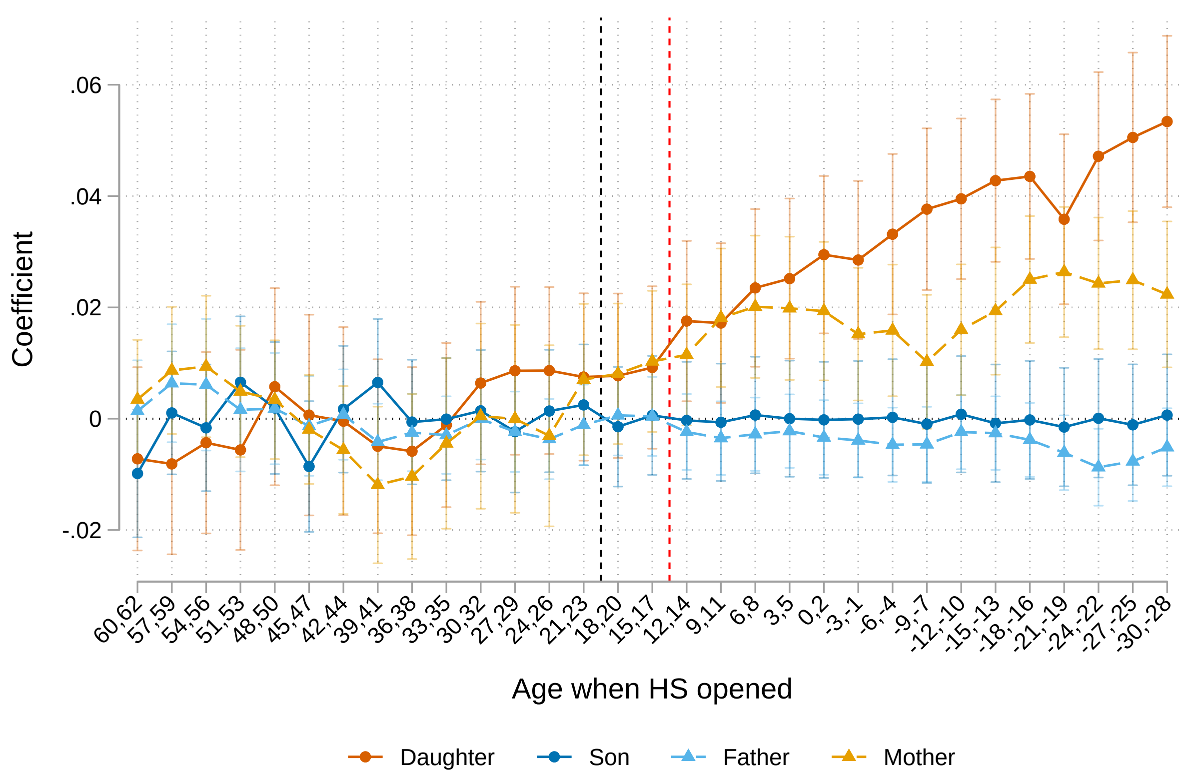
Notes: These panels show event-study estimates from Equation 2 for employment in the indicated occupations. See Figure 4 notes for sample and specification details.

Figure A10: Impact on age 30s LFP and occupational income, compared to impact on parents

(a) Log occupational income



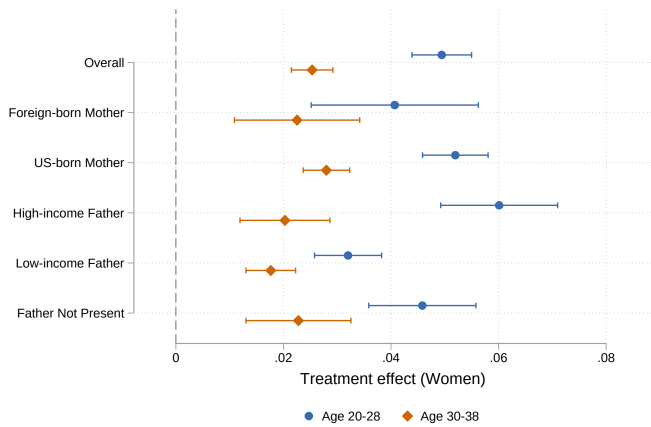
(b) Labor force participation



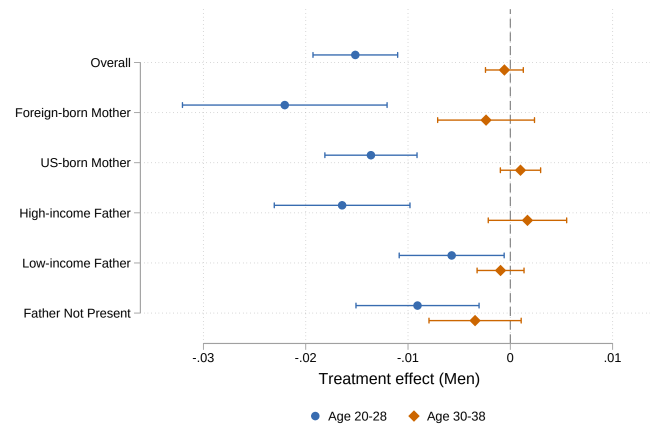
Notes: These panels show event-study estimates from Equation 2 separately for each indicated outcome and sample.

Figure A11: Heterogeneity by family background

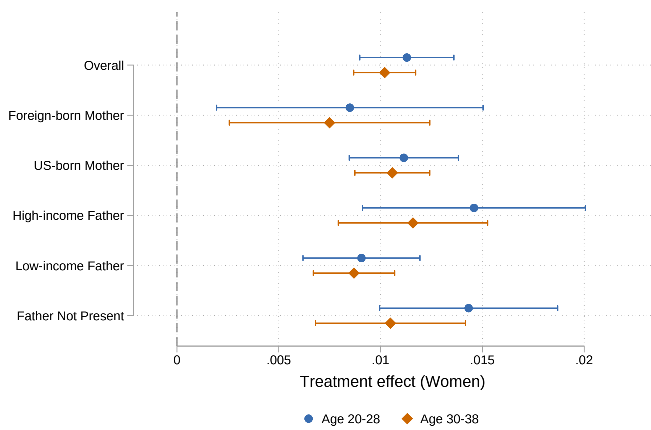
(a) Labor force participation, women



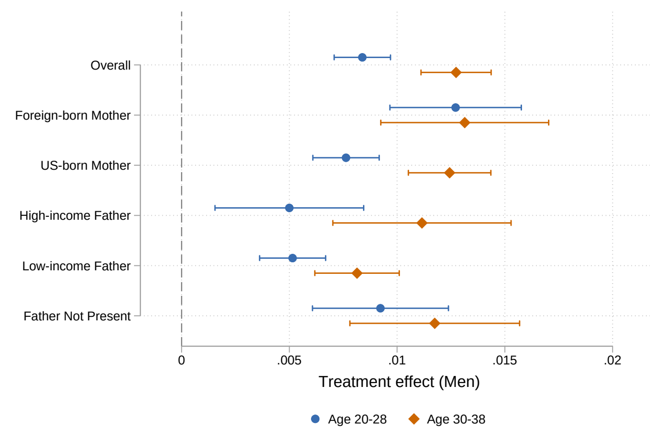
(b) Labor force participation, men



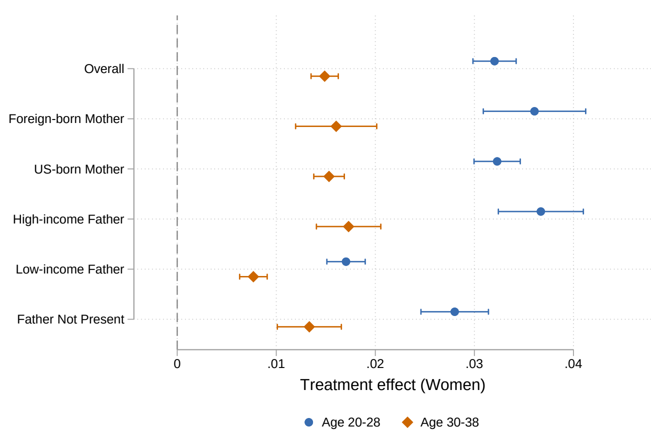
(c) Professional occupation, women



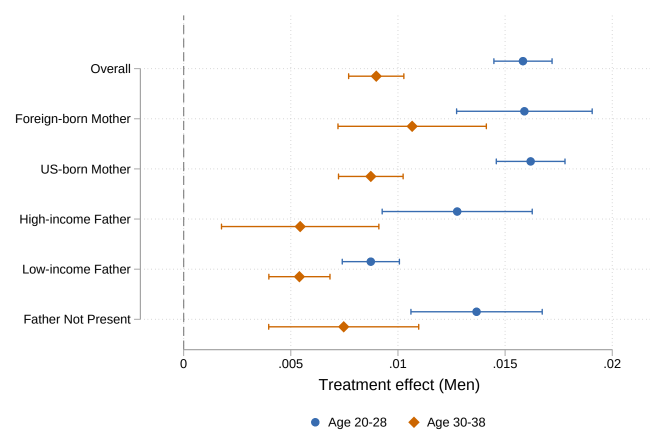
(d) Professional occupation, men



(e) Clerical occupation, women



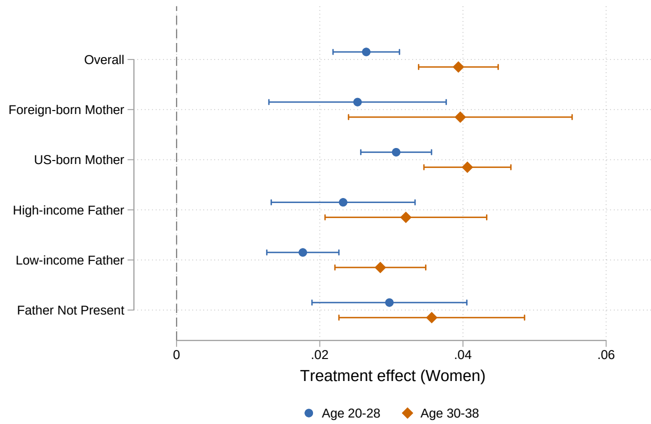
(f) Clerical occupation, men



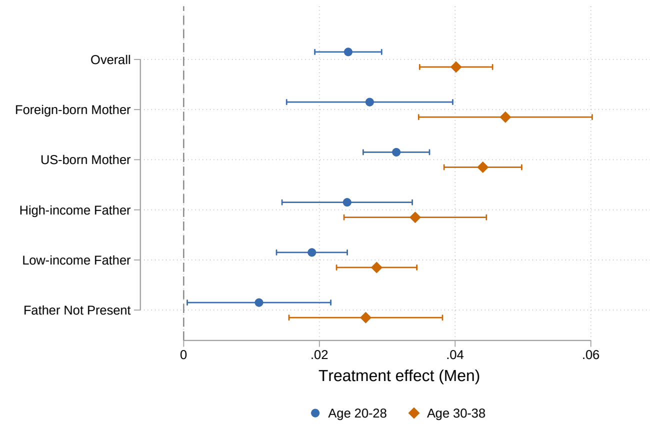
Notes: These panels show pre-post difference-in-differences estimates for key adult outcomes for people of different family backgrounds (outcomes are measured at age 20-28 and age 30-38). Each coefficient represents a separate regression.

Figure A12: Heterogeneity by family background, continued

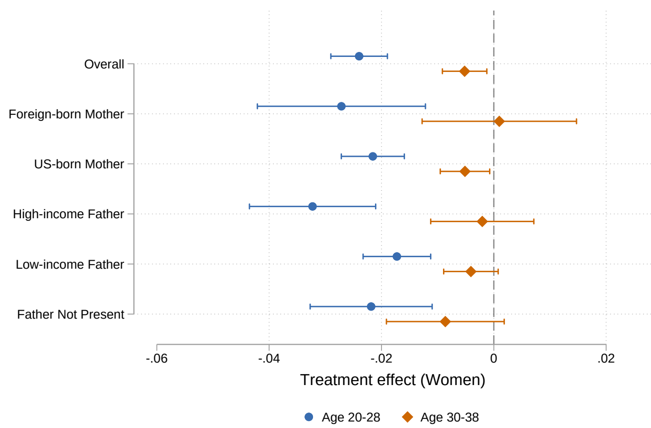
(a) Geographic mobility, women



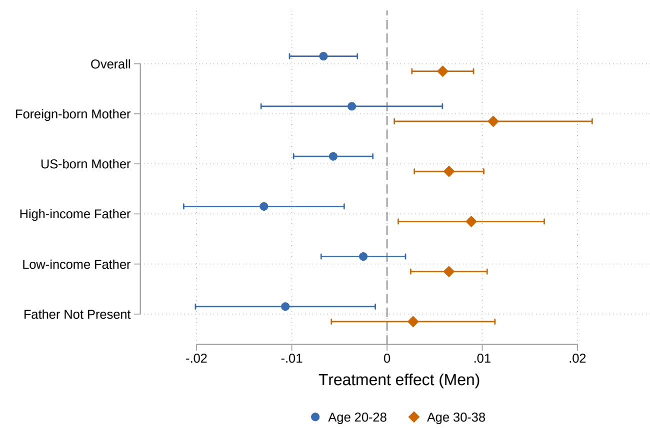
(b) Geographic mobility, men



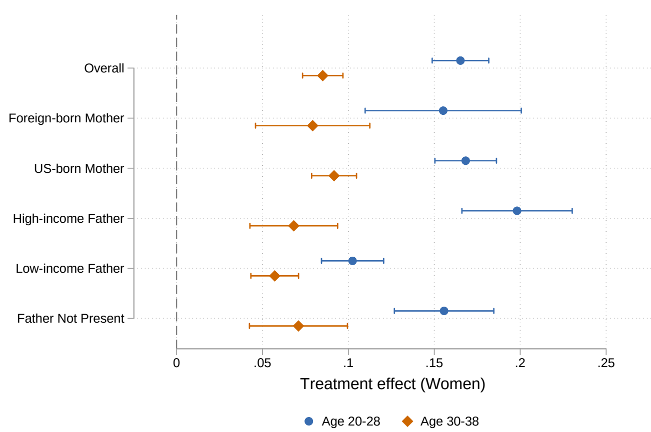
(c) Marriage, women



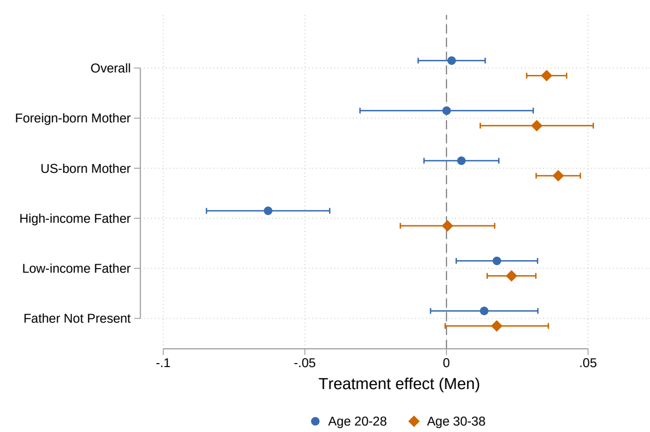
(d) Marriage, men



(e) Logged occupational income, women

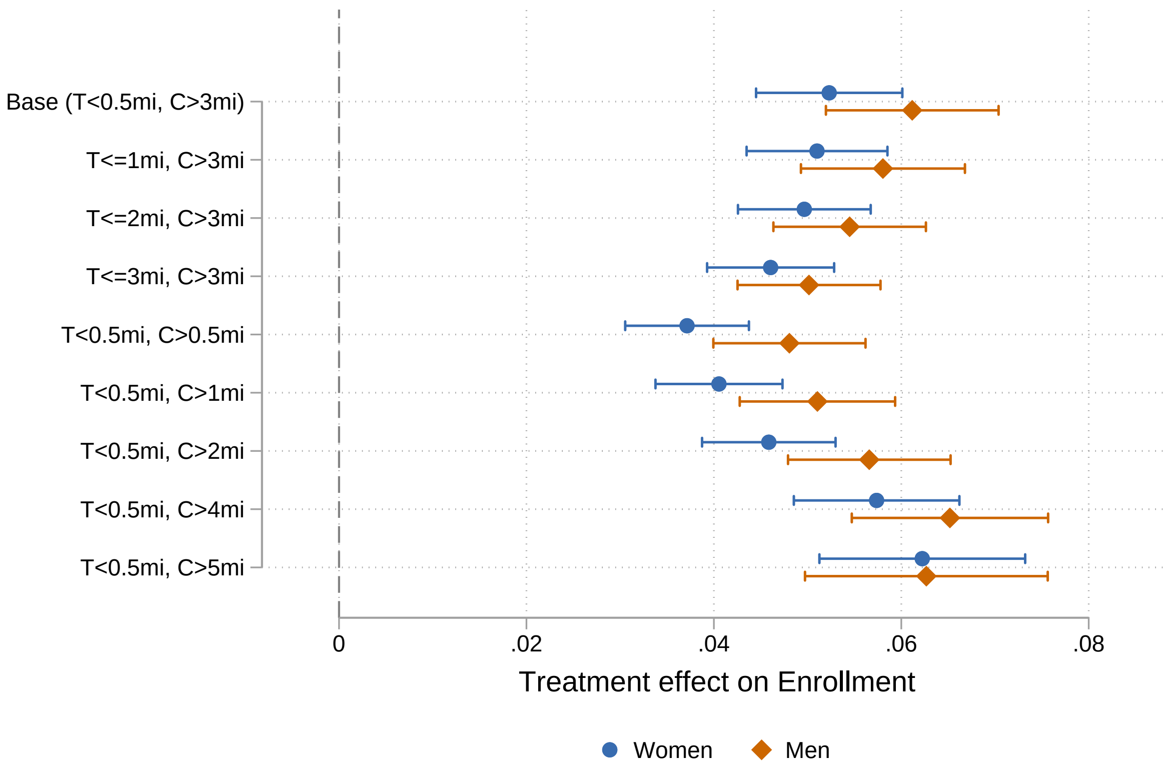


(f) Logged occupational income, men



Notes: These panels show pre-post difference-in-differences estimates for key adult outcomes for people of different family backgrounds (outcomes are measured at age 20-28 and age 30-38). Each coefficient represents a separate regression.

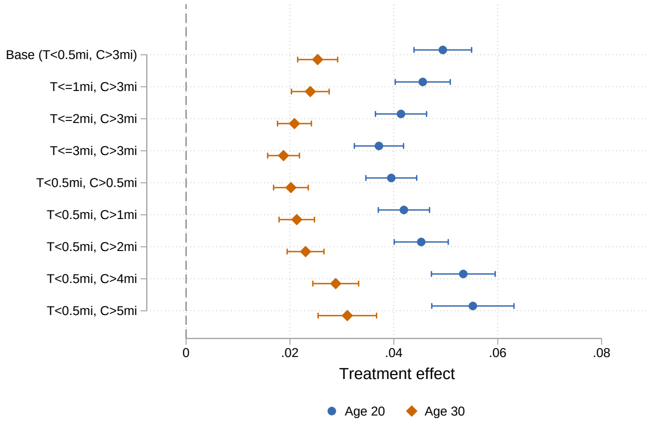
Figure A13: Distance attendance robustness, by gender



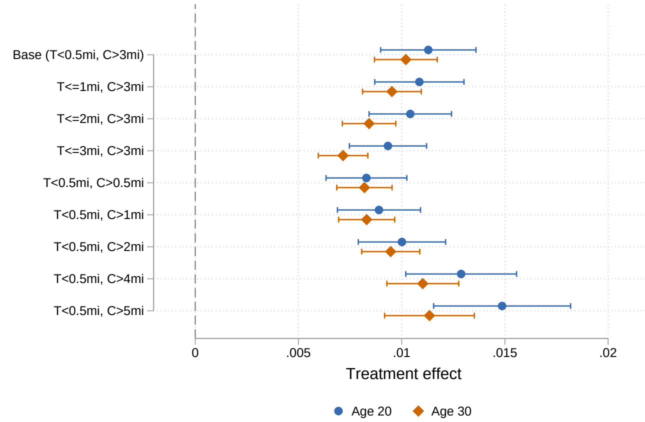
Notes: This figure shows pre-post difference-in-differences estimates for attendance (ages 17–18) under alternative treatment and control distance thresholds. Each coefficient represents a separate regression varying either the treatment radius (holding control at ≥3 miles) or the control radius (holding treatment at ≤0.5 miles). Standard errors clustered by nearest high school. Model includes city, birth-year, census year, county-by-birth-year, and high-school-opening-year-by-birth-year fixed effects.

Figure A14: Treatment distance thresholds long run robustness results, women

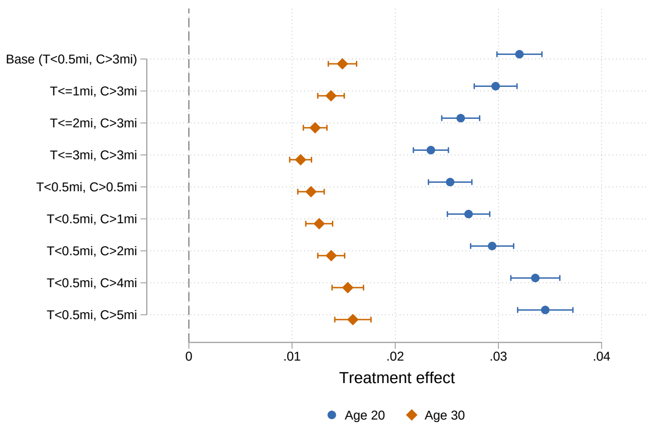
(a) Labor force participation



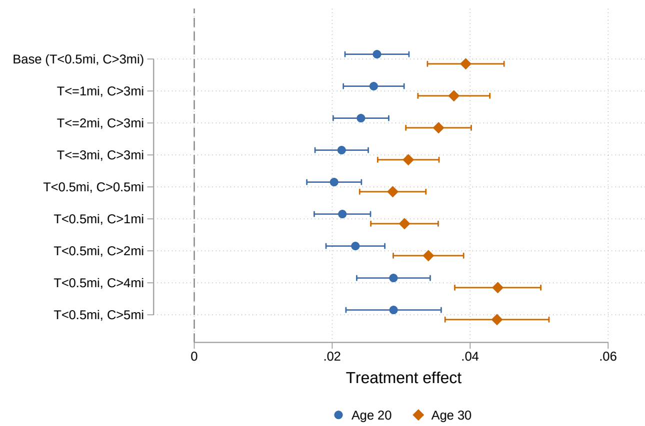
(b) Professional



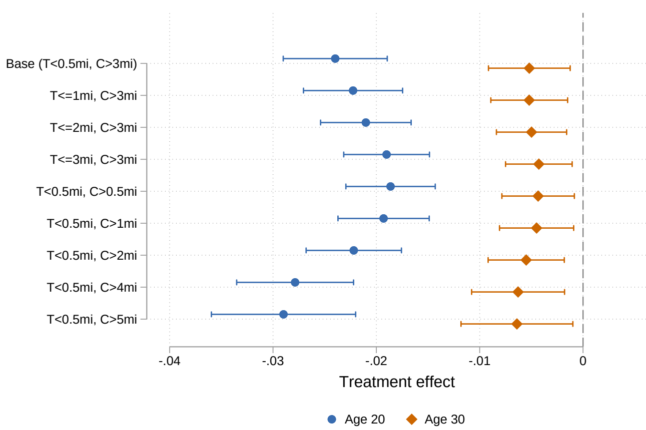
(c) Clerk



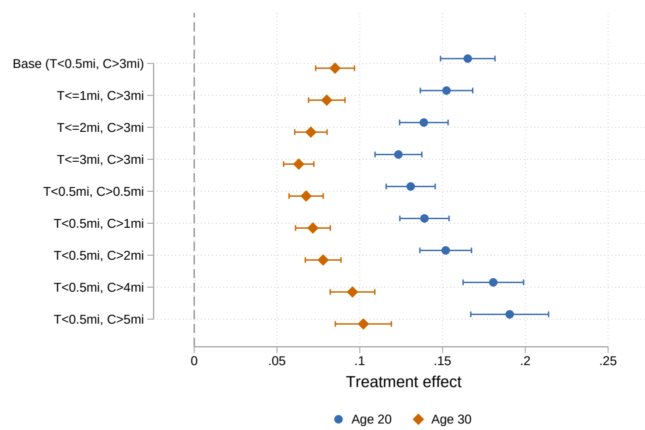
(d) Geographic mobility



(e) Marriage



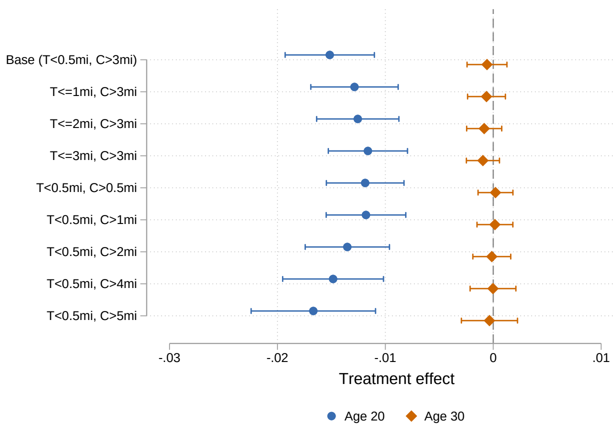
(f) Logged occupational income



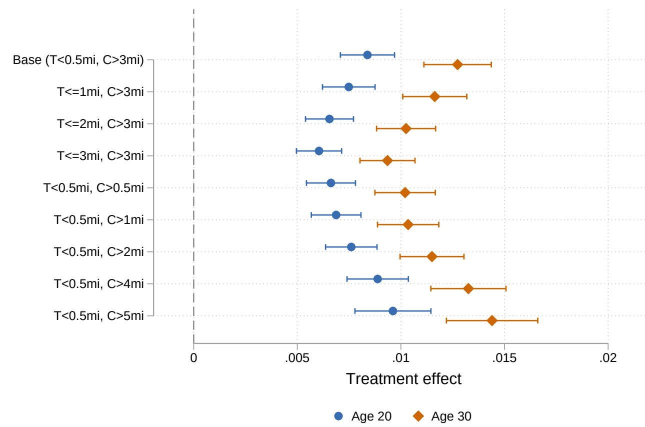
Notes: These panels show pre-post difference-in-differences estimates for six key adult outcomes (ages 20–28) under alternative treatment and control distance thresholds. Each coefficient represents a separate regression. See Figure A13 notes for specification details.

Figure A15: Treatment distance thresholds long run robustness results, men

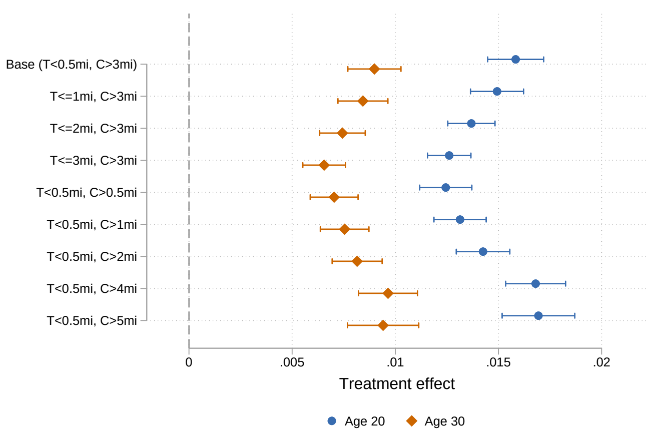
(a) Labor force participation



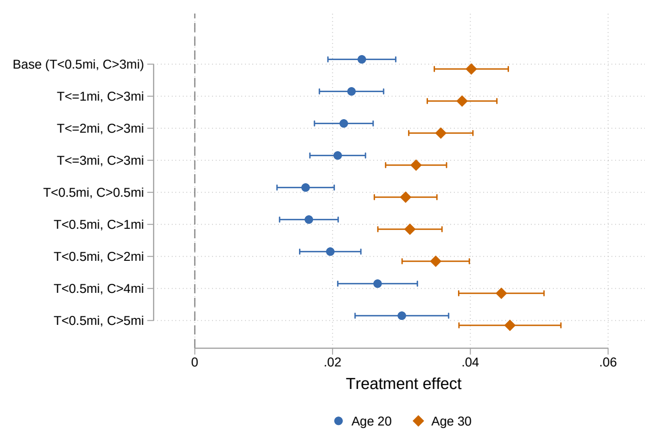
(b) Professional



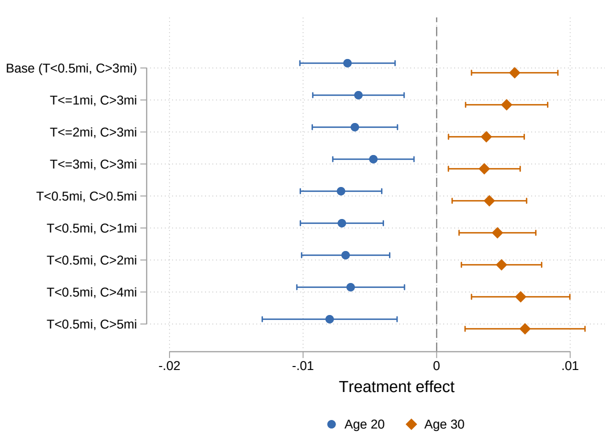
(c) Clerk



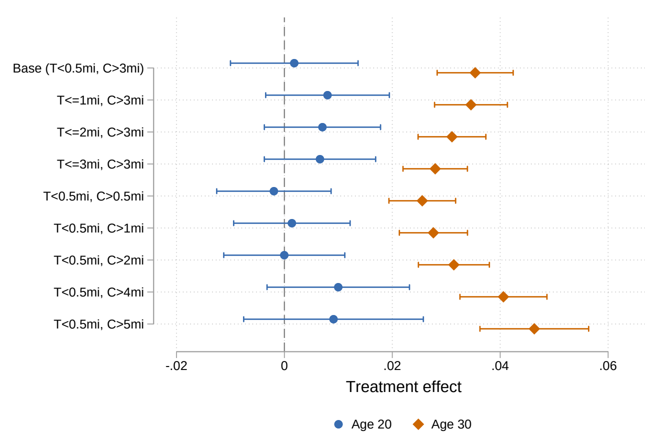
(d) Geographic mobility



(e) Marriage

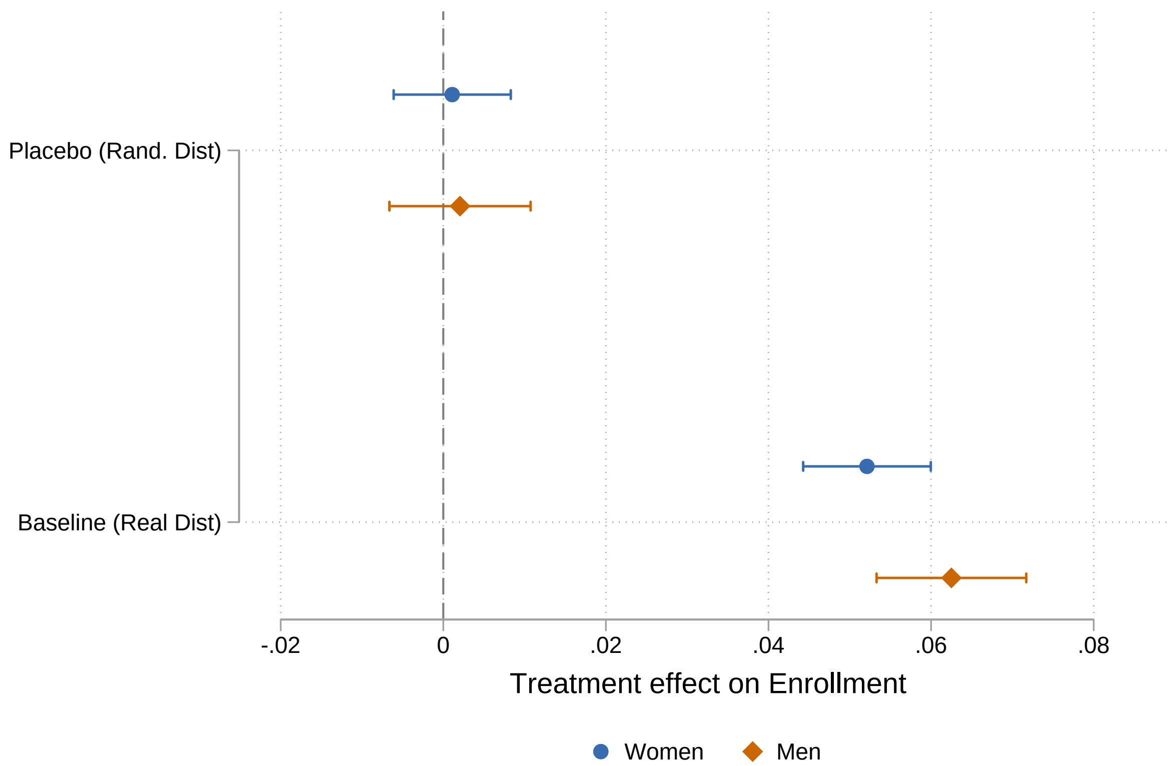


(f) Logged occupational income



Notes: These panels show pre-post difference-in-differences estimates for six key adult outcomes (ages 20–28) under alternative treatment and control distance thresholds. Each coefficient represents a separate regression. See Figure A13 notes for specification details.

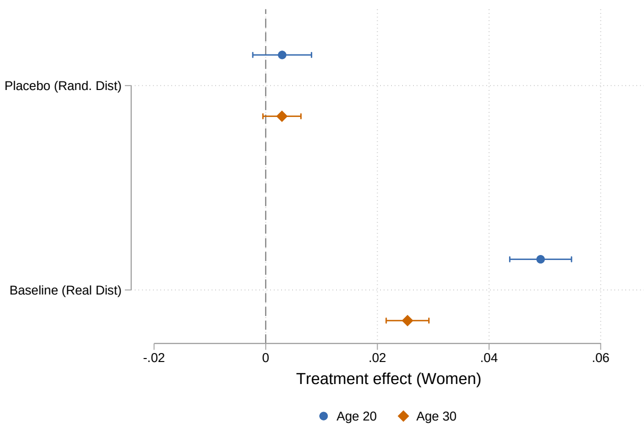
Figure A16: Randomized distance: Attendance, by gender



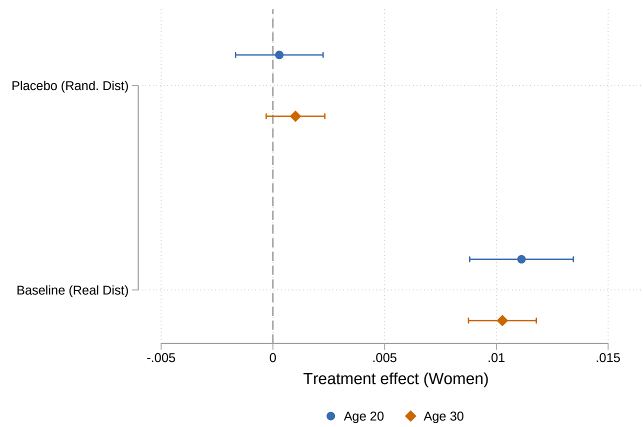
Notes: These panels show placebo tests where each city's actual distance to the nearest high school is replaced with a distance drawn without replacement from the distribution of actual distances in our sample. Top panels show results using randomized distances; bottom panels show baseline results using actual distances for comparison. See Figure 2 notes for specification details.

Figure A17: Randomized distance: long run female outcomes

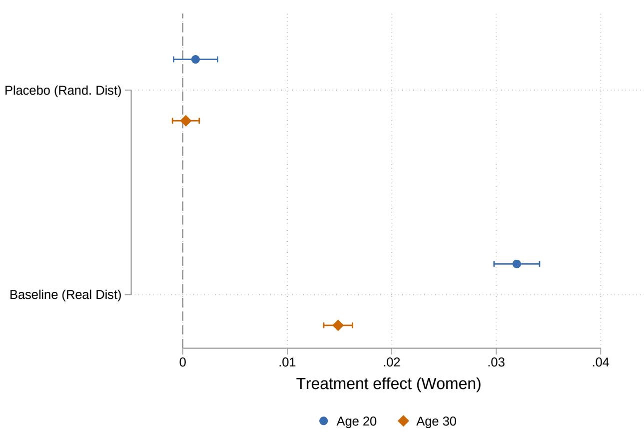
(a) Labor force participation



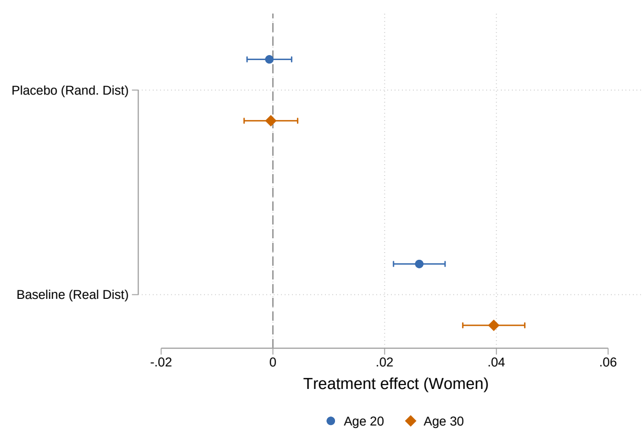
(b) Professional



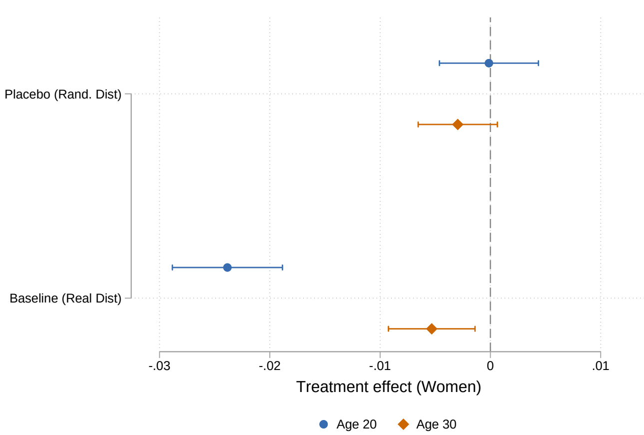
(c) Clerk



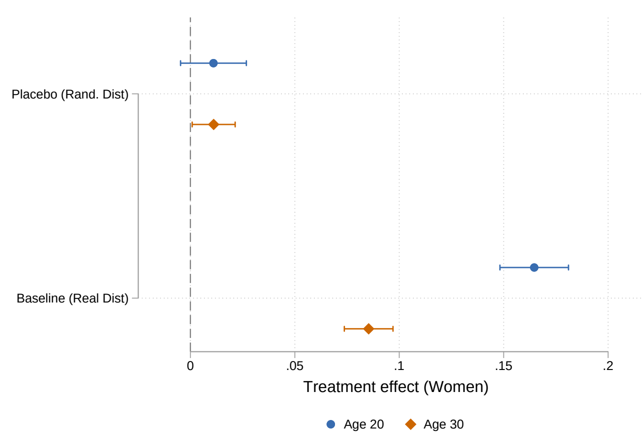
(d) Geographic mobility



(e) Marriage



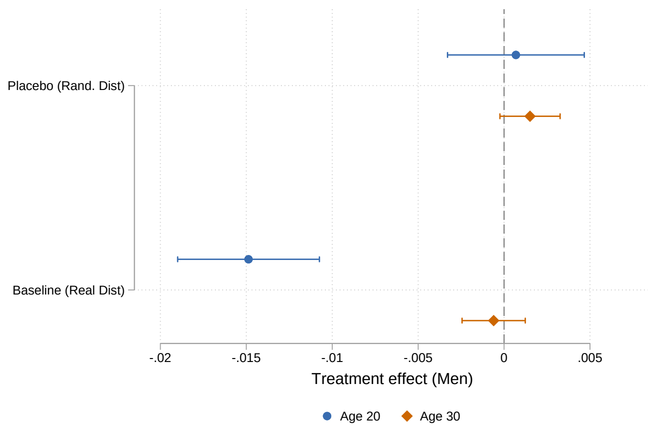
(f) Logged occupational income



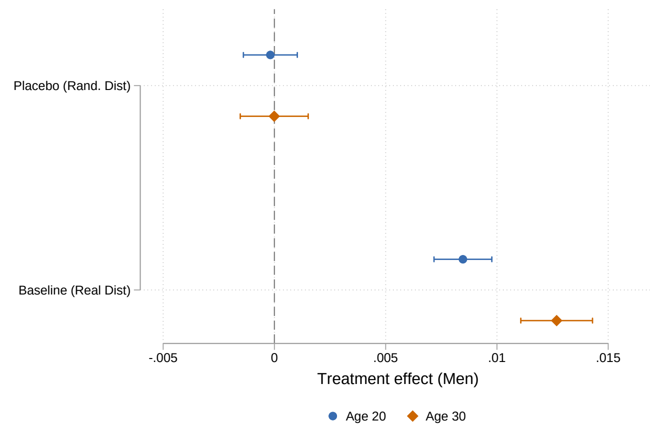
Notes: These panels show placebo tests where each city's actual distance to the nearest high school is replaced with a distance drawn without replacement from the distribution of actual distances in our sample. Top panels show results using randomized distances; bottom panels show baseline results using actual distances for comparison. See Figure 4 notes for specification details.

Figure A18: Randomized distance: long run male outcomes

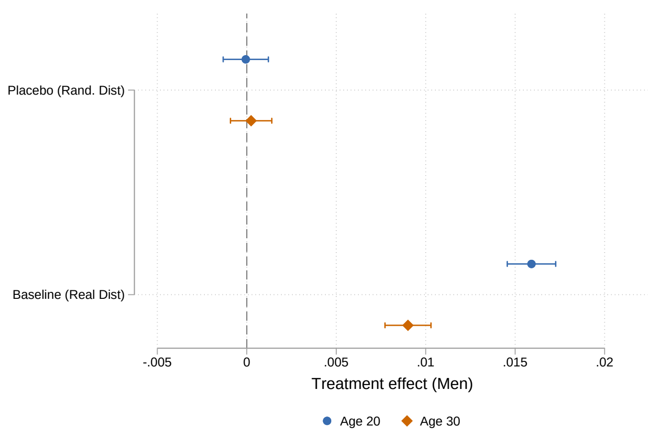
(a) Labor force participation



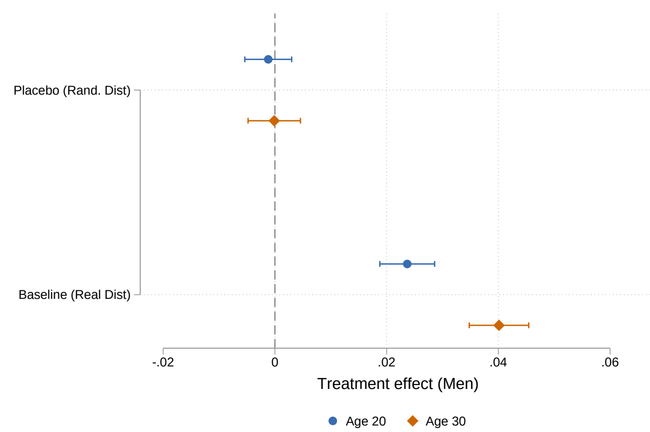
(b) Professional



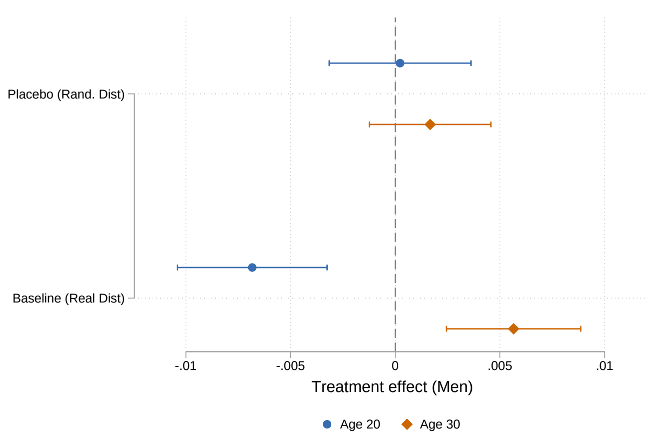
(c) Clerk



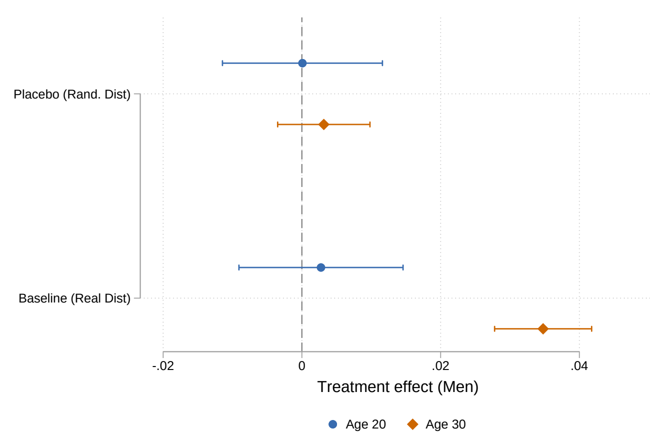
(d) Geographic mobility



(e) Marriage

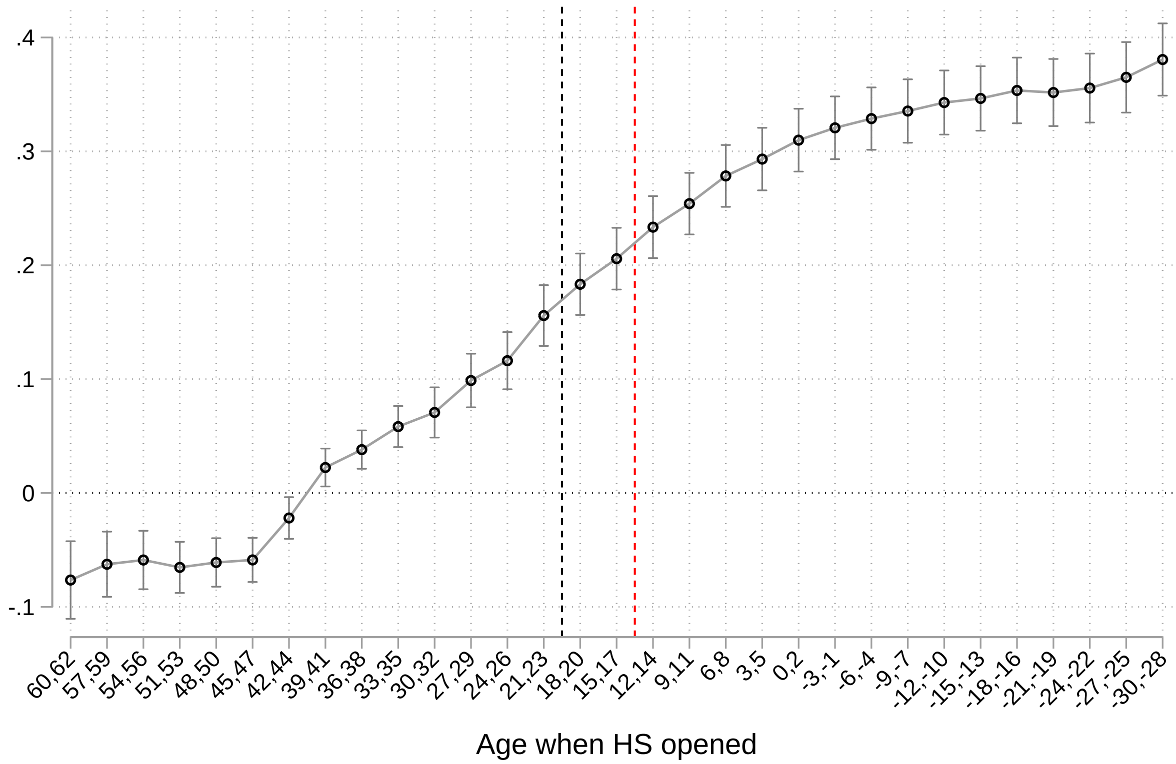


(f) Logged occupational income



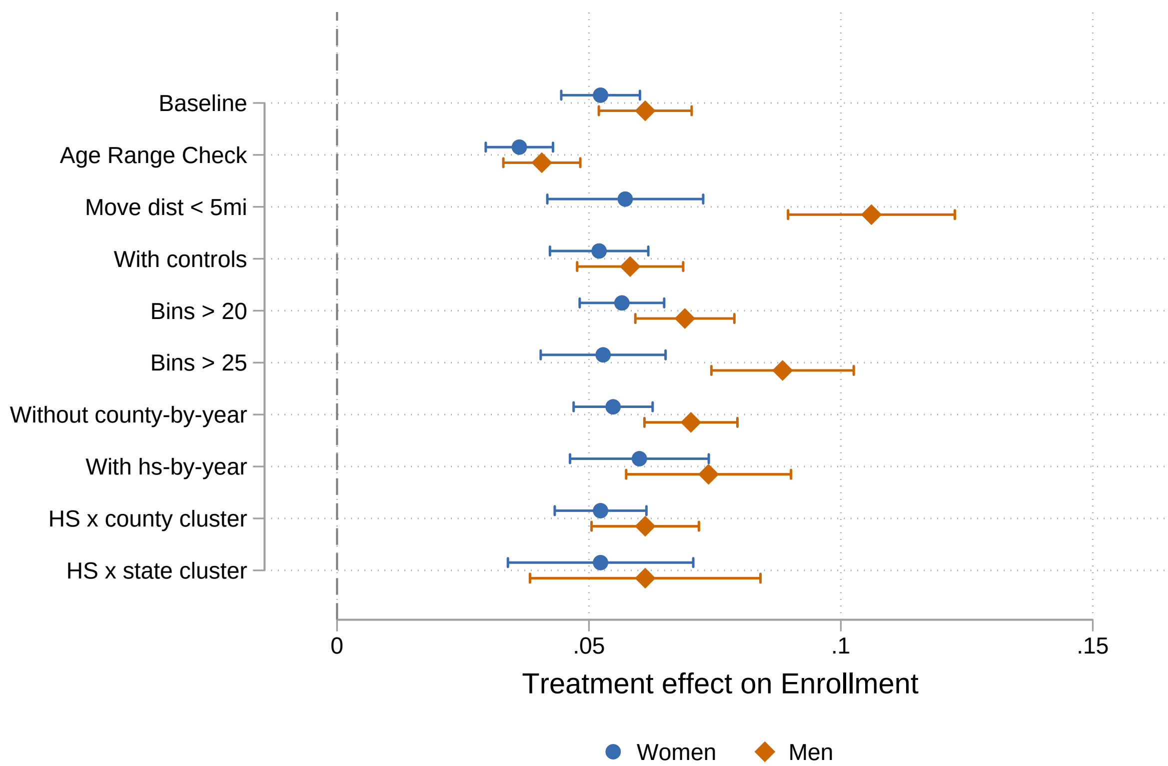
Notes: These panels show placebo tests where each city's actual distance to the nearest high school is replaced with a distance drawn without replacement from the distribution of actual distances in our sample. Top panels show results using randomized distances; bottom panels show baseline results using actual distances for comparison. See Figure 4 notes for specification details.

Figure A19: Railroads entry as a function of high school formation dates



Notes: This figure shows event-study estimates from Equation 1 with an indicator for railroad access as the outcome variable. The x-axis represents age when the nearest high school opened. The black dashed line indicates partially treated cohorts; the red dashed line indicates children who could have attended all four traditional years of high school. Treatment group includes children living ≤ 0.5 miles from a high school; control group lives ≥ 3 miles away. The smooth increase across cohorts shows that while cities with high schools were positively selected on railroad access, this selection does not change sharply at the time of high school entry.

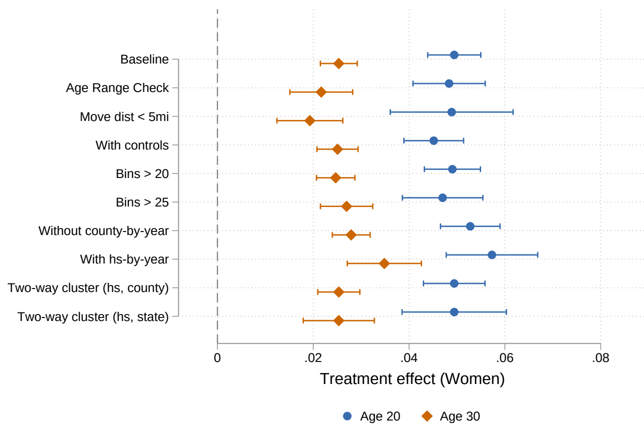
Figure A20: Attendance specification robustness summary, by gender



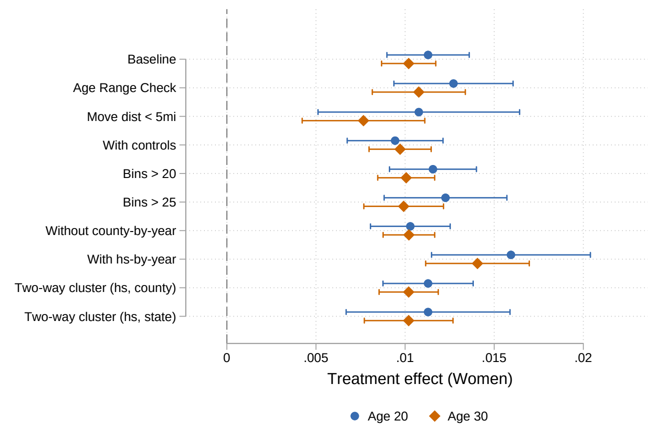
Notes: This figure summarizes pre-post difference-in-differences estimates for attendance (ages 17–18) under alternative specifications and sample restrictions. Each coefficient represents a separate regression. Robustness checks include: expanding the age range to 14–18; excluding county-by-birth-year fixed effects; including high-school-by-birth-year fixed effects; restricting to non-movers (individuals who moved <5 miles between ages 0–8 and 10–18); requiring at least 20 or 25 age bins for panel balance; conditioning on parental occupation scores, parental labor force participation, and railroad access; and two-way clustering by nearest high school and county or state. See Figure 2 notes for baseline specification details.

Figure A21: Specification robustness: long run female outcomes

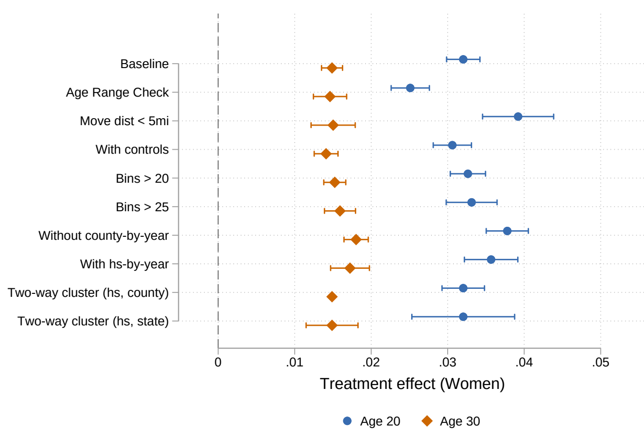
(a) Labor force participation



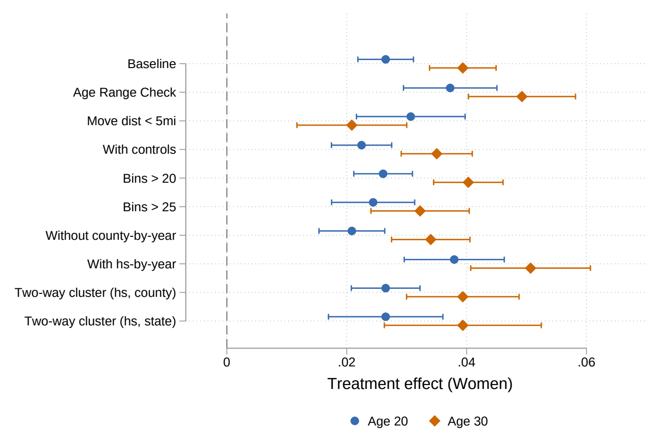
(b) Professional



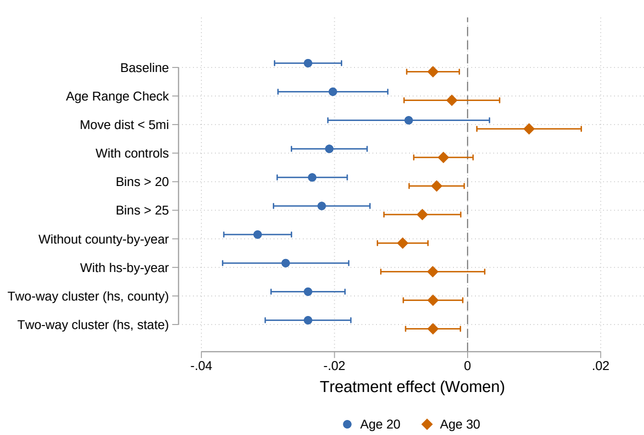
(c) Clerk



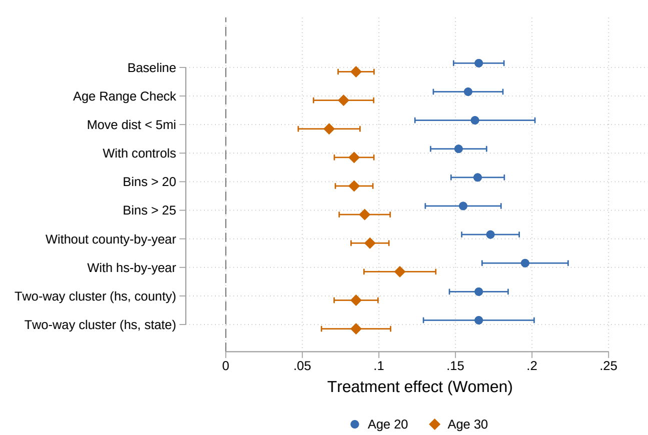
(d) Geographic mobility



(e) Marriage



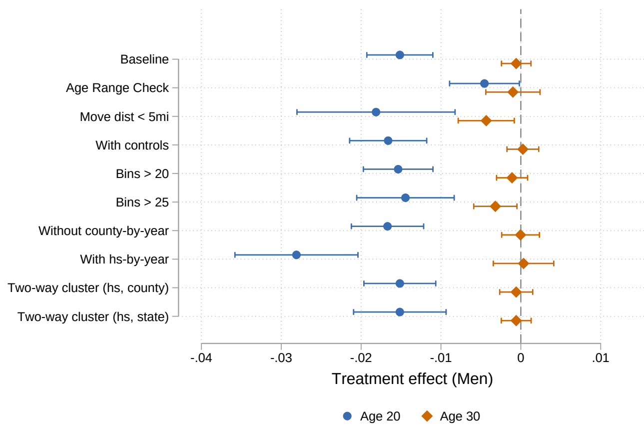
(f) Logged occupational income



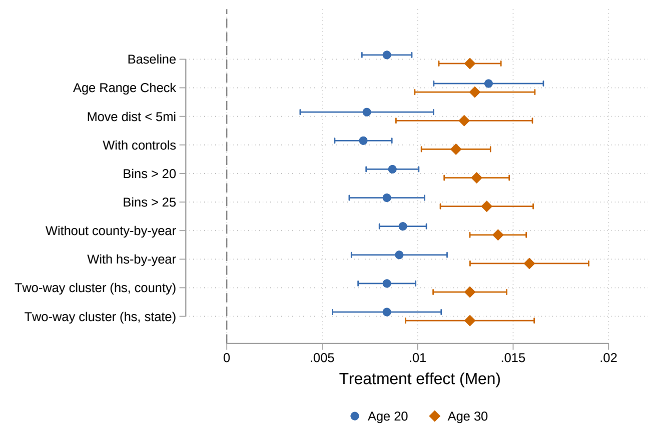
Notes: This figure summarizes pre-post difference-in-differences estimates for six key adult outcomes (ages 20–28) for women. See Figure 4 notes for baseline specification details and Figure A20 for robustness test details.

Figure A22: Specification robustness: long run male outcomes

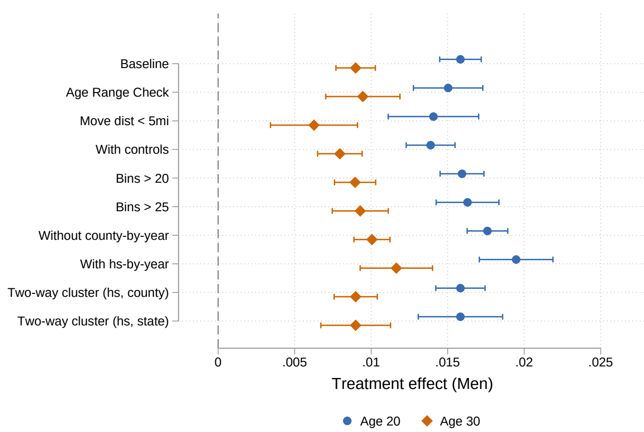
(a) Labor force participation



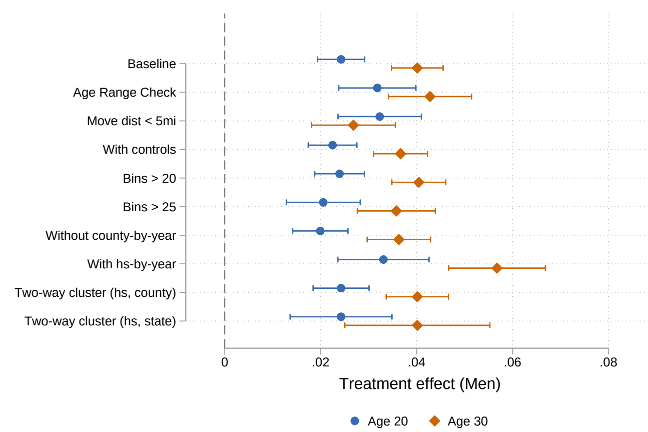
(b) Professional



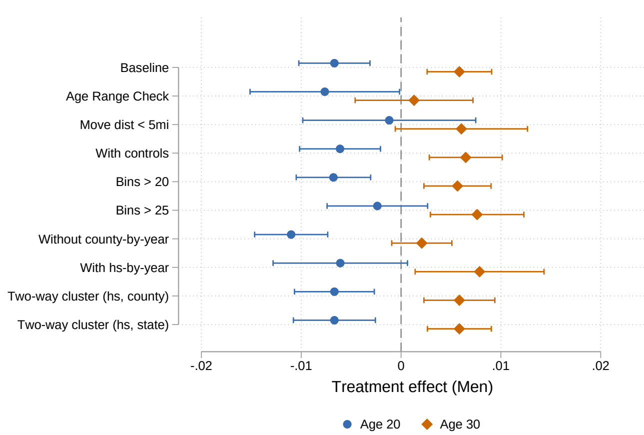
(c) Clerk



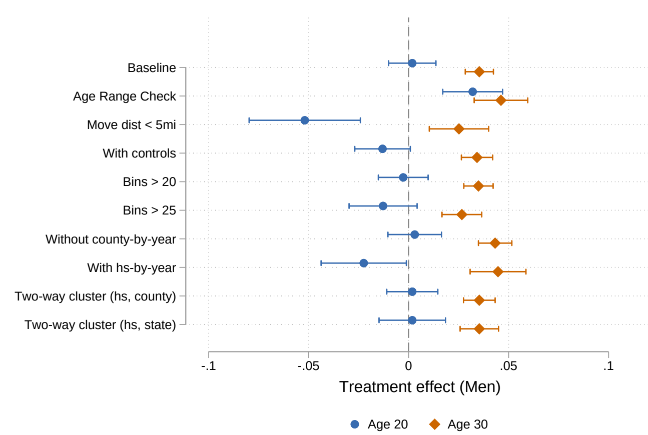
(d) Geographic mobility



(e) Marriage



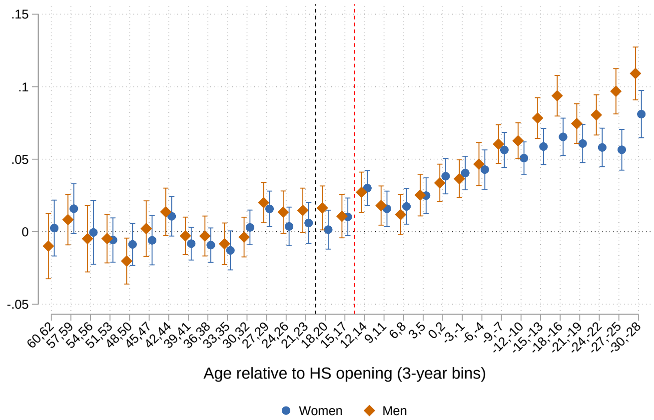
(f) Logged occupational income



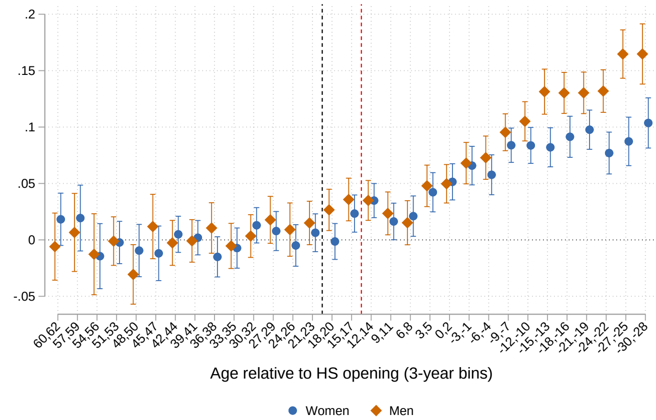
Notes: This figure summarizes pre-post difference-in-differences estimates for six key adult outcomes (ages 20–28) for men. See Figure 4 notes for baseline specification details and Figure A20 for robustness test details.

Figure A23: Attendance event-study robustness, by gender

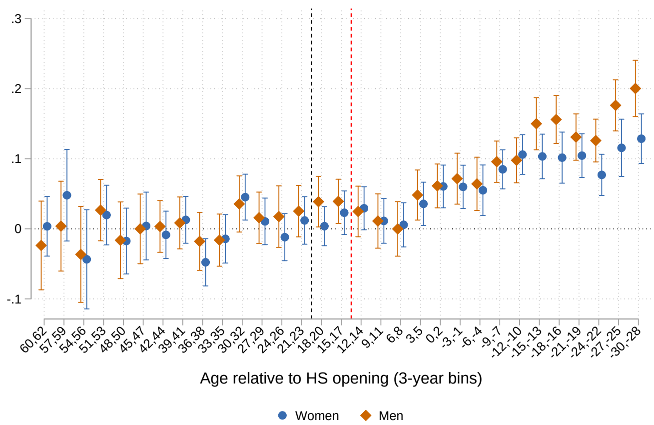
(a) 14-18 year old age range



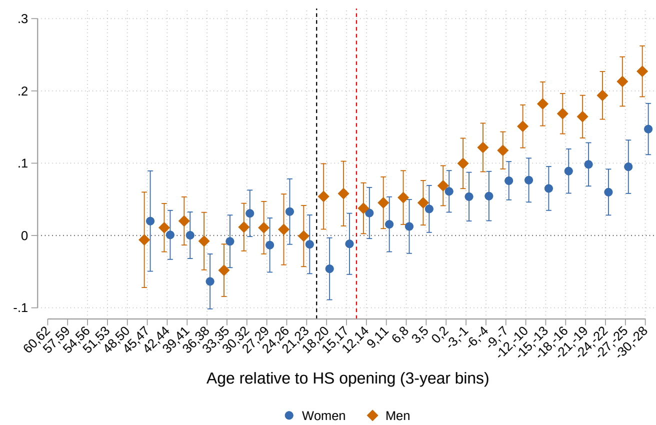
(b) Without county-by-year FEs



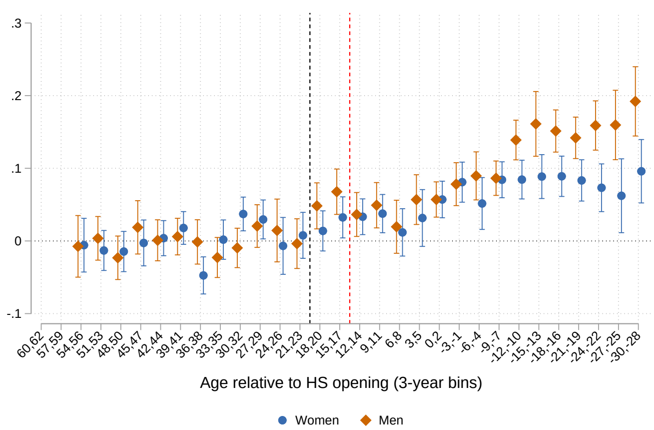
(c) Include HS-by-year FEs



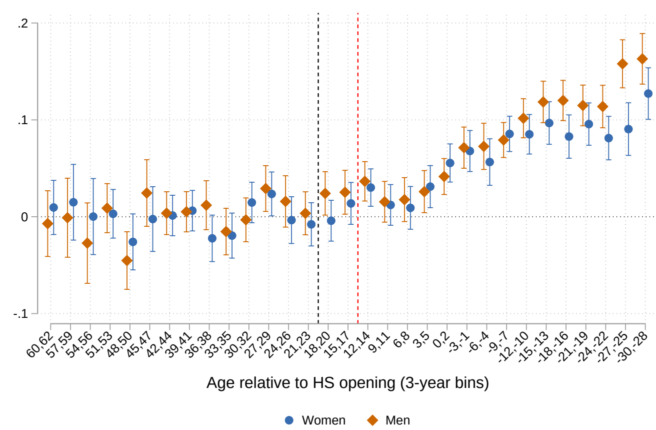
(d) Non-movers



(e) More balanced city panel



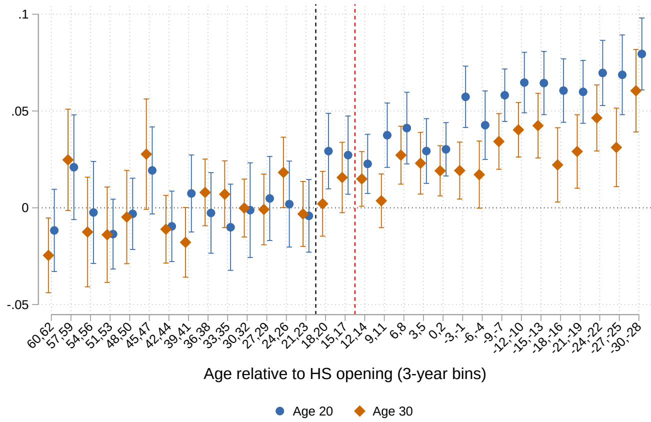
(f) With parental occupation and RR controls



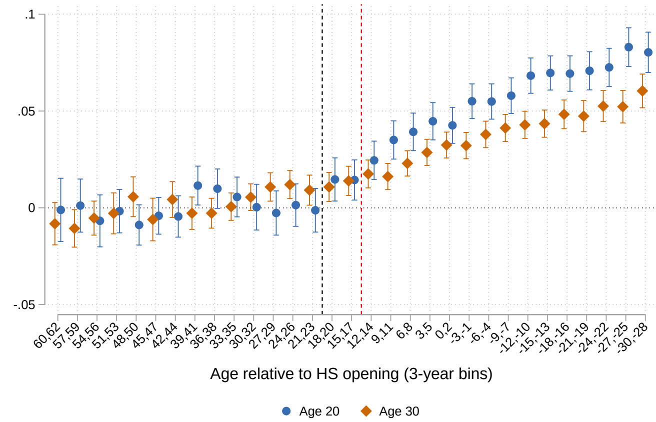
Notes: This figure shows our event study results for each indicated robustness check.

Figure A24: Labor Force Participation event-study robustness, women

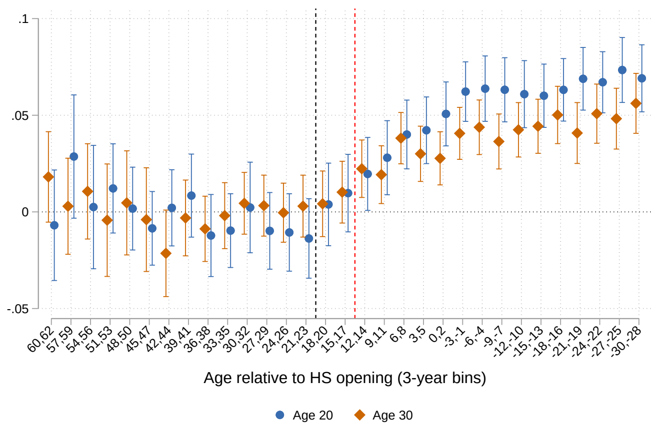
(a) Only link 17-18 year olds



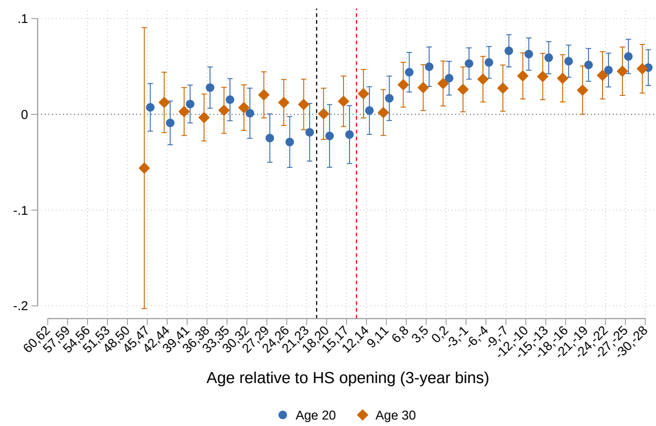
(b) Without county-by-year FEs



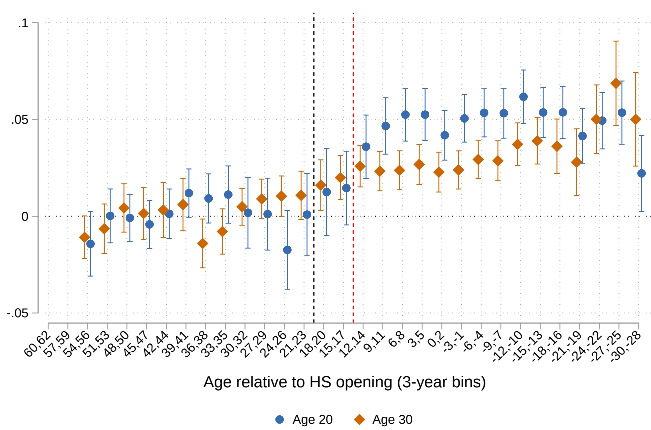
(c) Include HS-by-year FEs



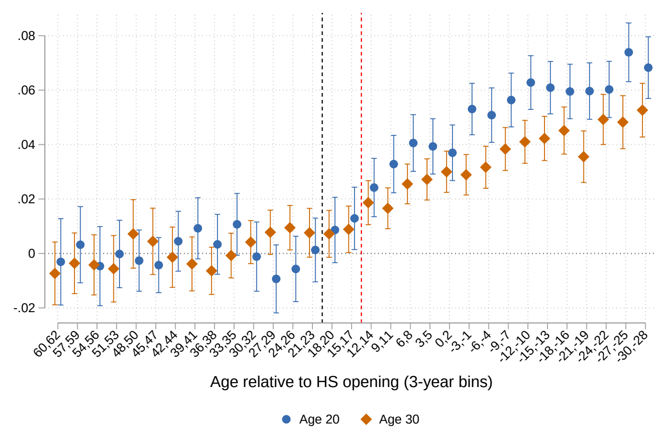
(d) Non-movers



(e) More balanced city panel



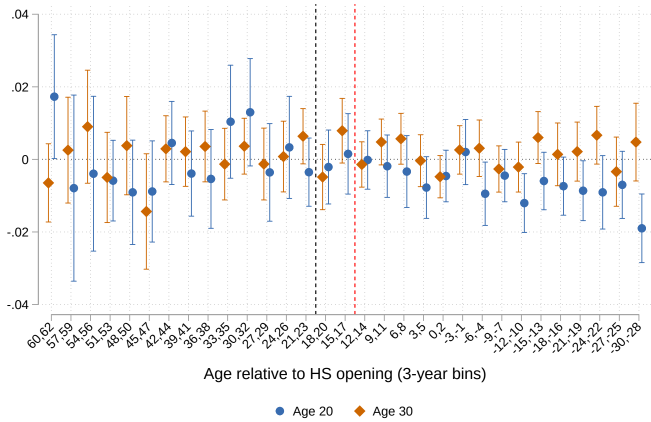
(f) With parental occupation and RR controls



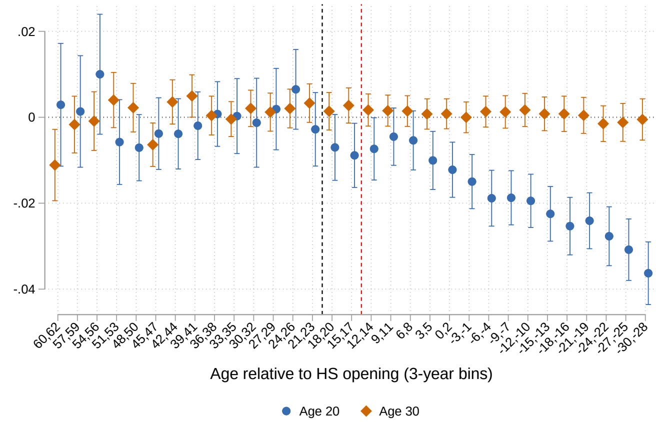
Notes: This figure shows our event study results for each indicated robustness check the indicated age groups. See Equation 2 for specification and sample details.

Figure A25: Labor Force Participation event-study robustness, men

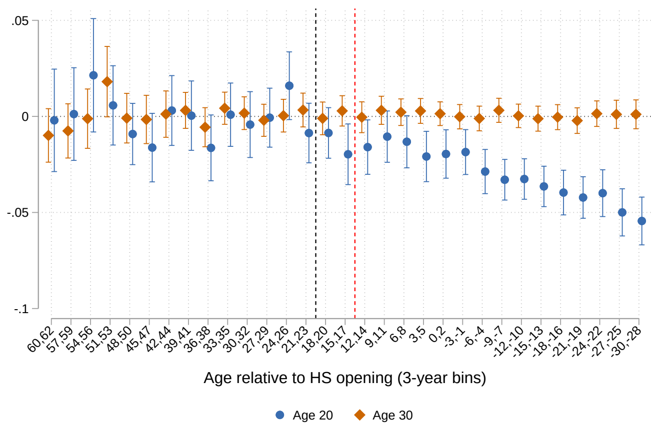
(a) Only link 17-18 year olds



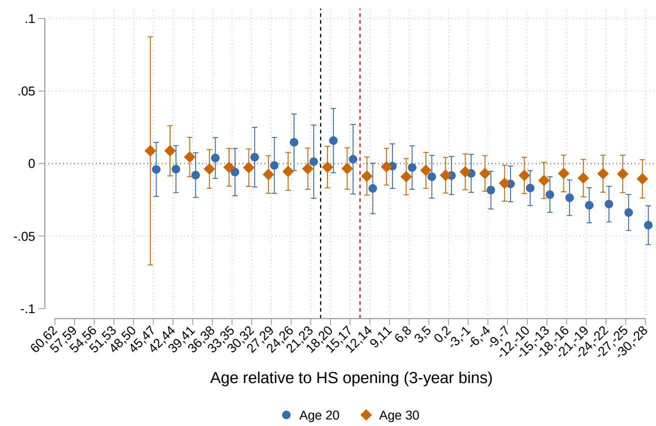
(b) Without county-by-year FEs



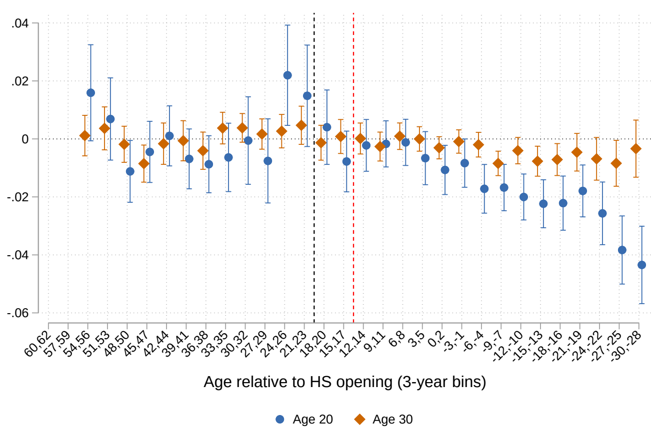
(c) Include HS-by-year FEs



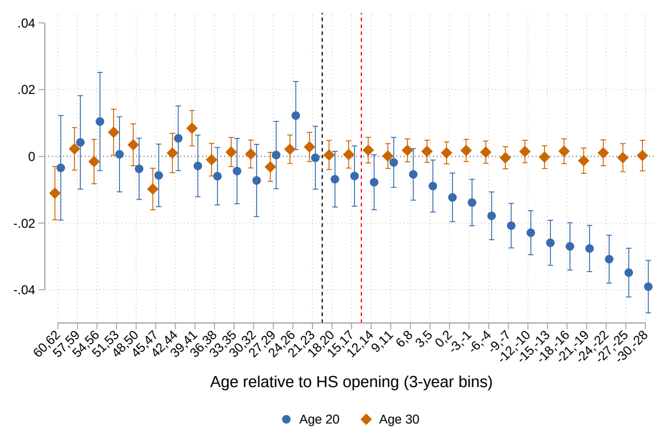
(d) Non-movers



(e) More balanced city panel



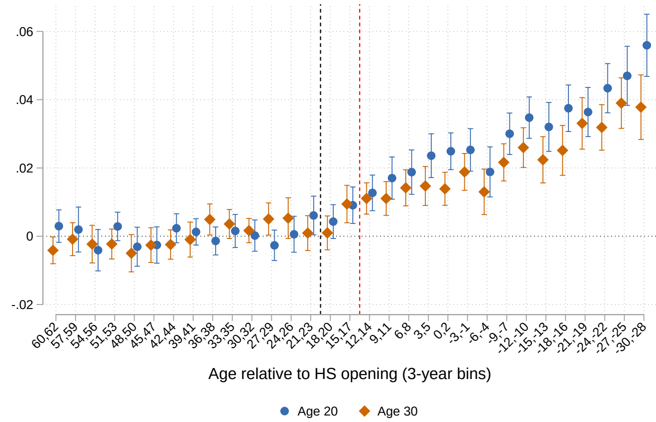
(f) With parental occupation and RR controls



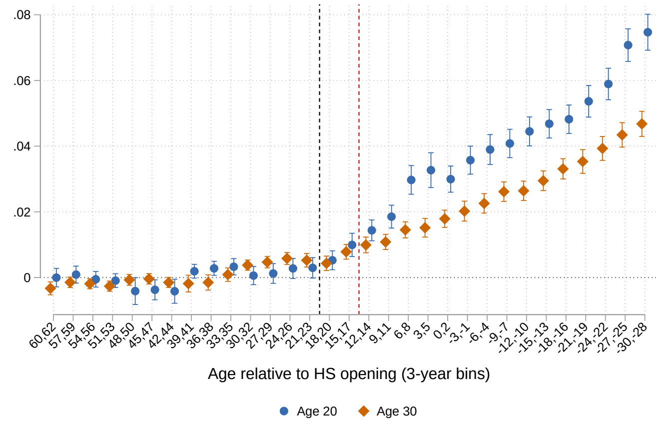
Notes: This figure shows our event study results for each indicated robustness check the indicated age groups. See Equation 2 for specification and sample details.

Figure A26: Clerk occupation event-study robustness, women

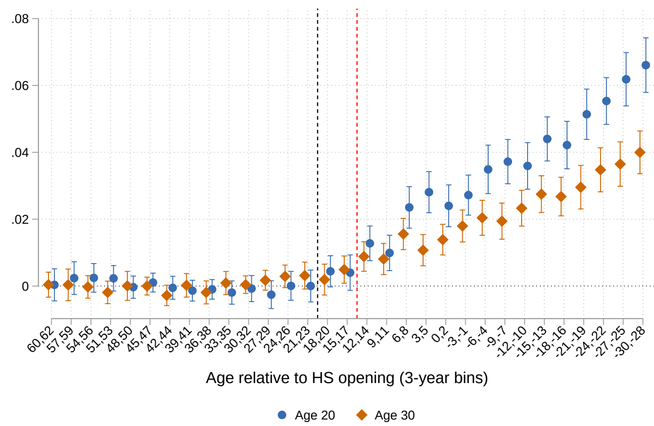
(a) Only link 17-18 year olds



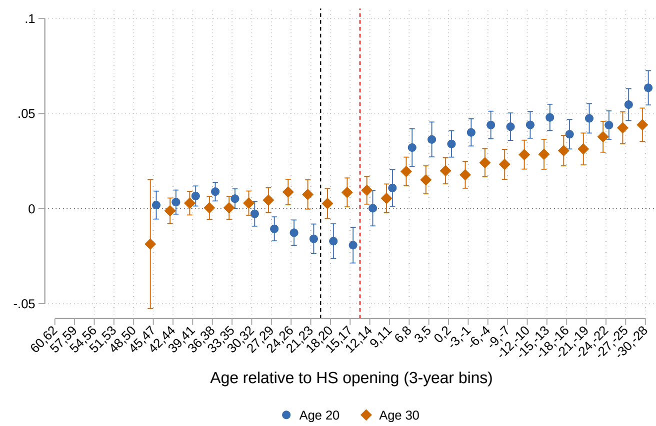
(b) Without county-by-year FEs



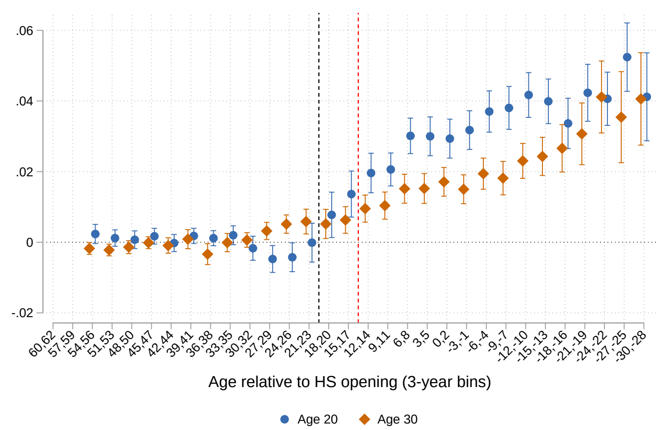
(c) Include HS-by-year FEs



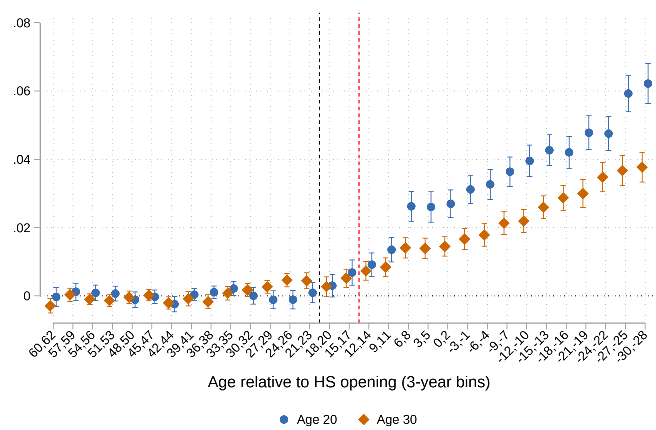
(d) Non-movers



(e) More balanced city panel



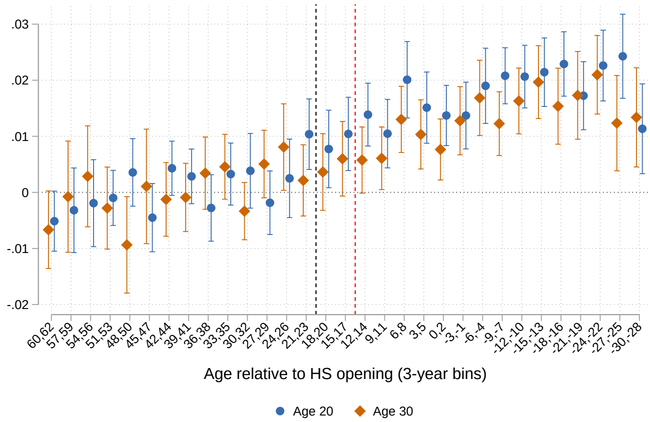
(f) With parental occupation and RR controls



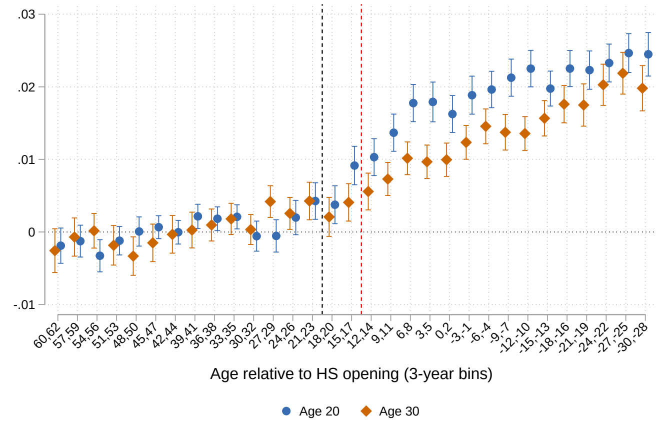
Notes: This figure shows our event study results for each indicated robustness check the indicated age groups. See Equation 2 for specification and sample details.

Figure A27: Clerk occupation event-study robustness, men

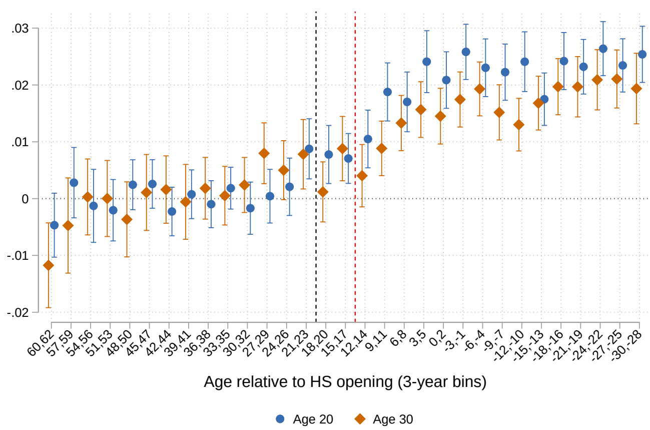
(a) Only link 17-18 year olds



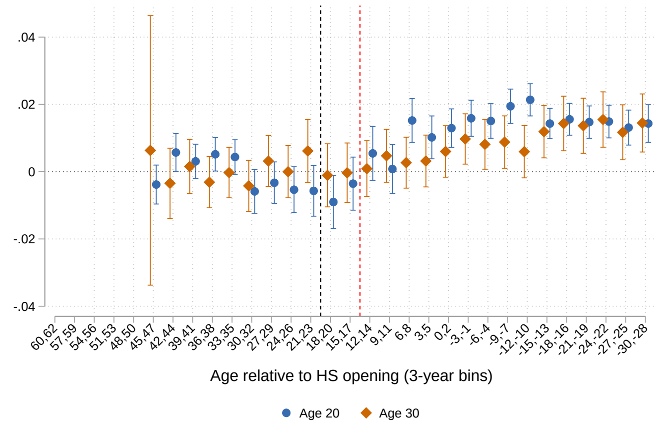
(b) Without county-by-year FEs



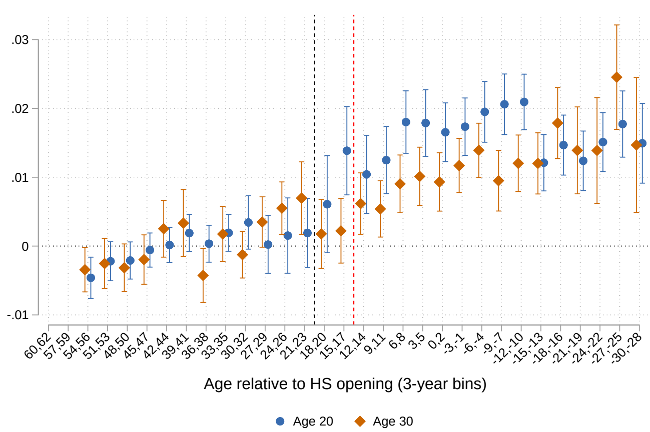
(c) Include HS-by-year FEs



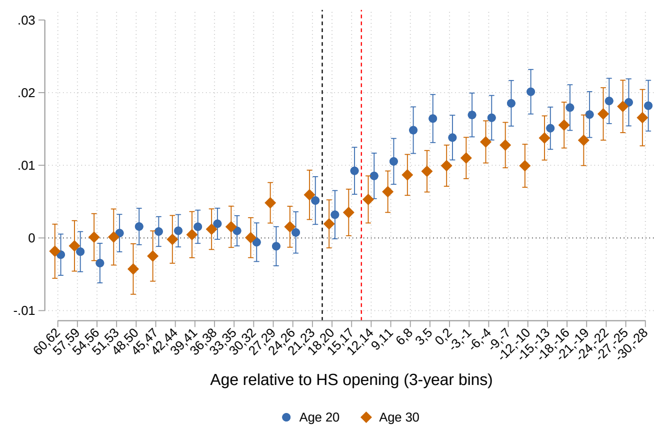
(d) Non-movers



(e) More balanced city panel

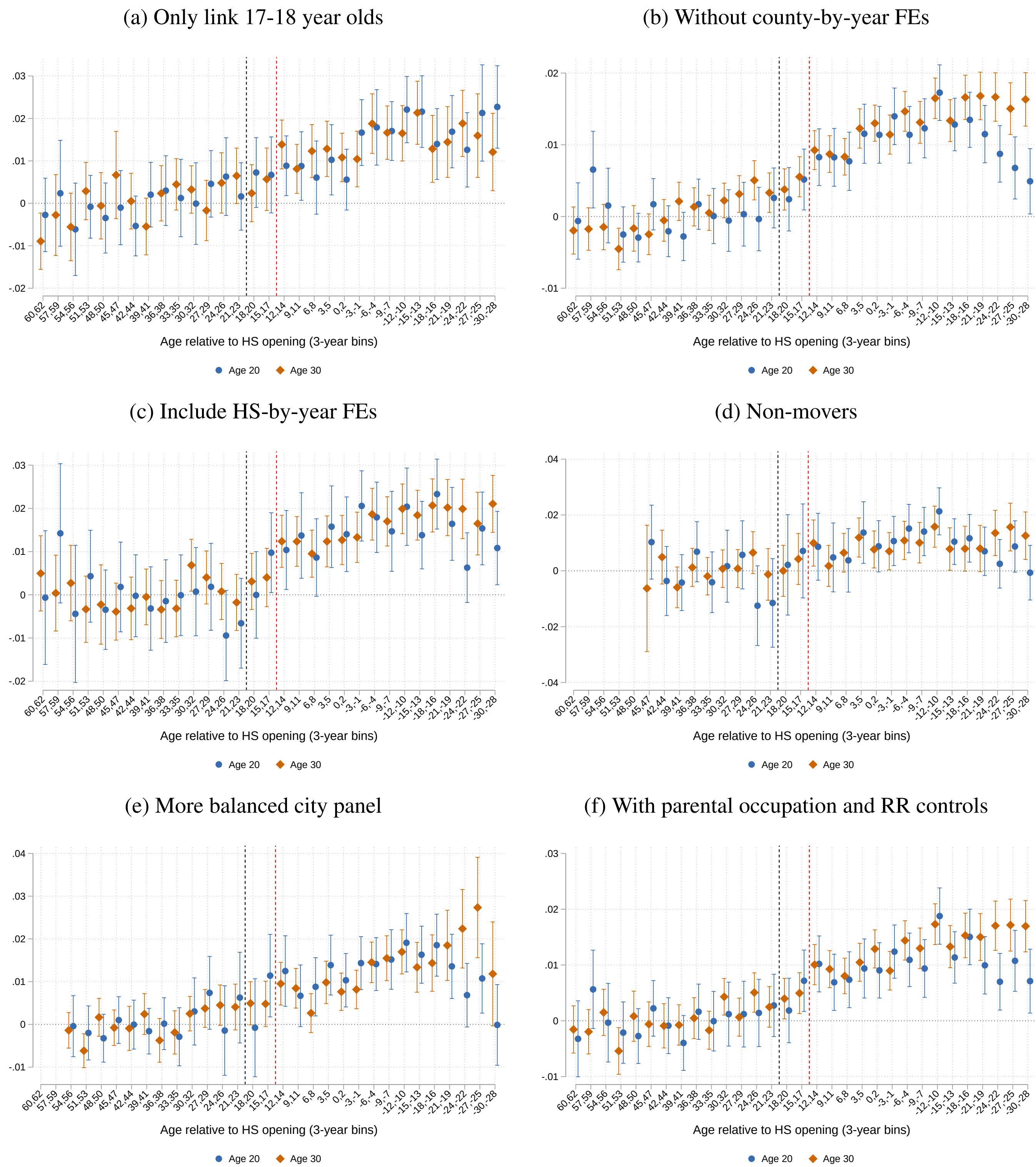


(f) With parental occupation and RR controls



Notes: This figure shows our event study results for each indicated robustness check the indicated age groups. See Equation 2 for specification and sample details.

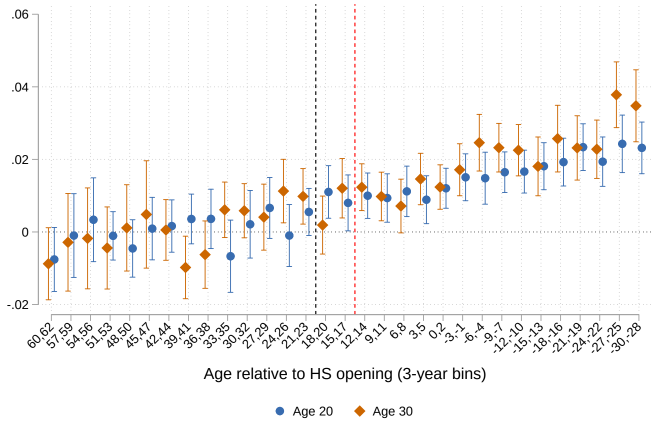
Figure A28: Professional occupation event-study robustness, women



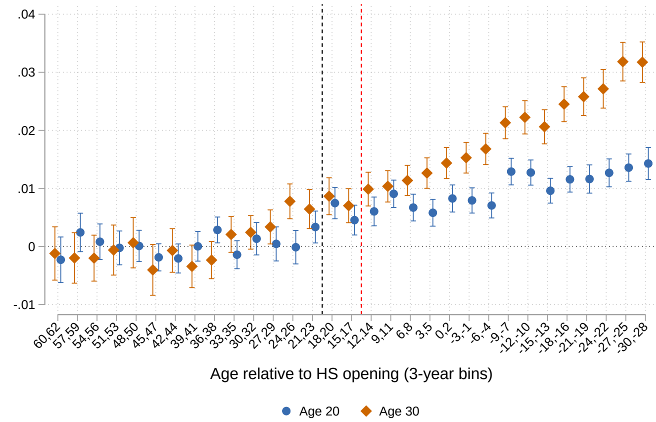
Notes: This figure shows our event study results for each indicated robustness check the indicated age groups. See Equation 2 for specification and sample details.

Figure A29: Professional occupation event-study robustness, men

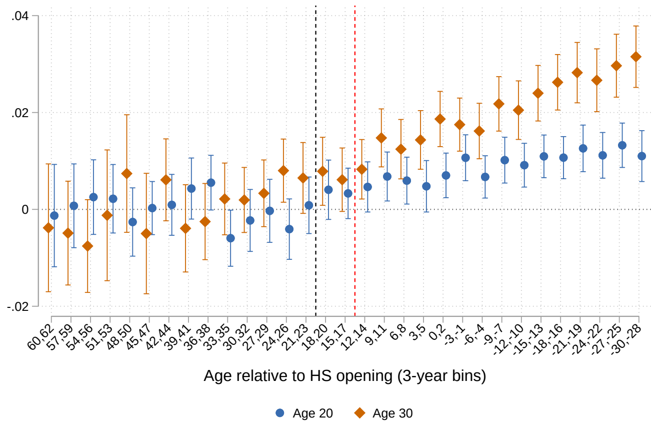
(a) Only link 17-18 year olds



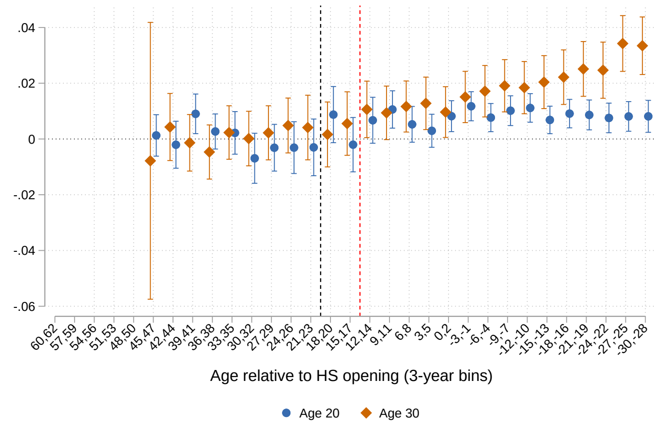
(b) Without county-by-year FEs



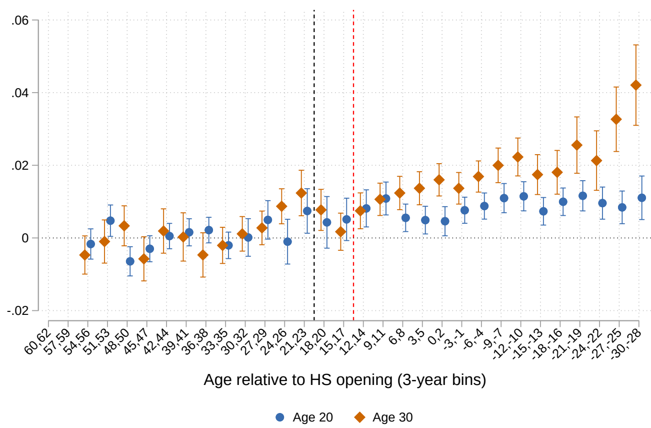
(c) Include HS-by-year FEs



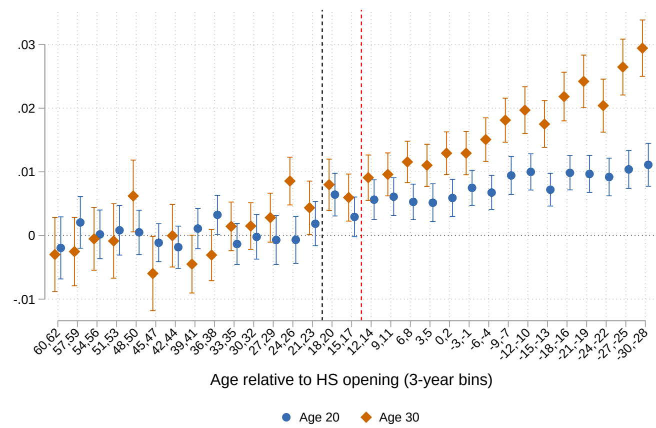
(d) Non-movers



(e) More balanced city panel



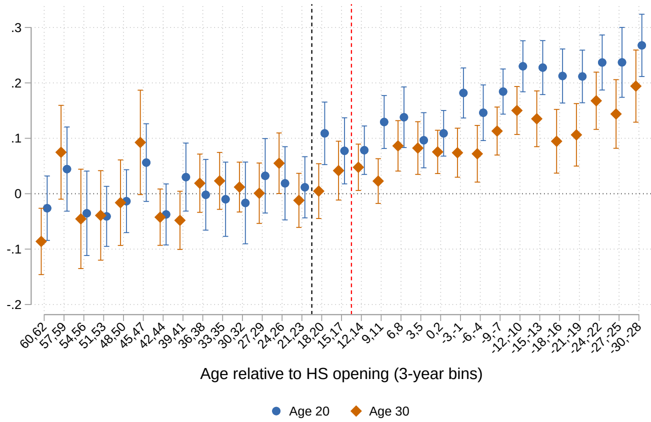
(f) With parental occupation and RR controls



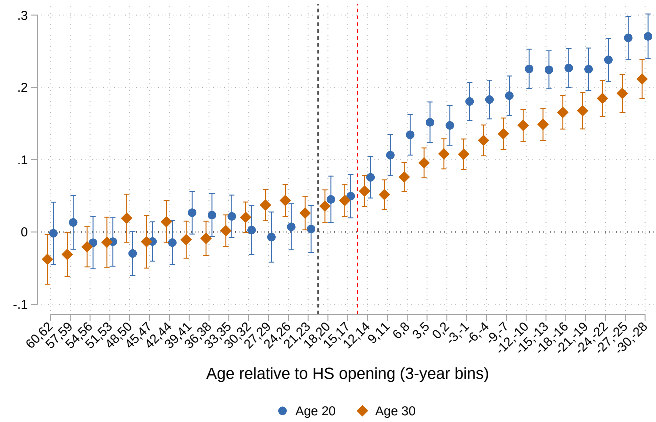
Notes: This figure shows our event study results for each indicated robustness check the indicated age groups. See Equation 2 for specification and sample details.

Figure A30: Log occupational income robustness, women

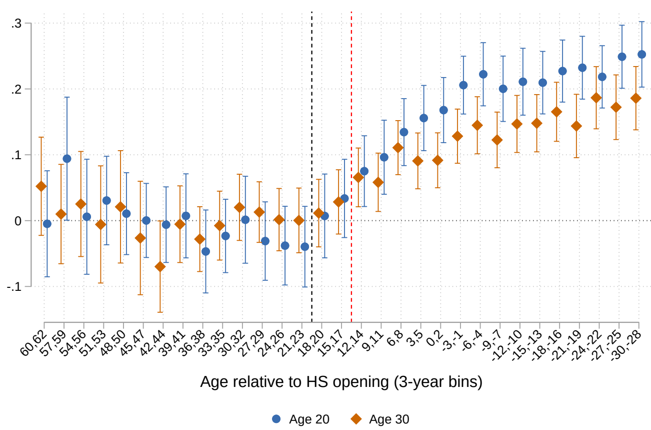
(a) Only link 17-18 year olds



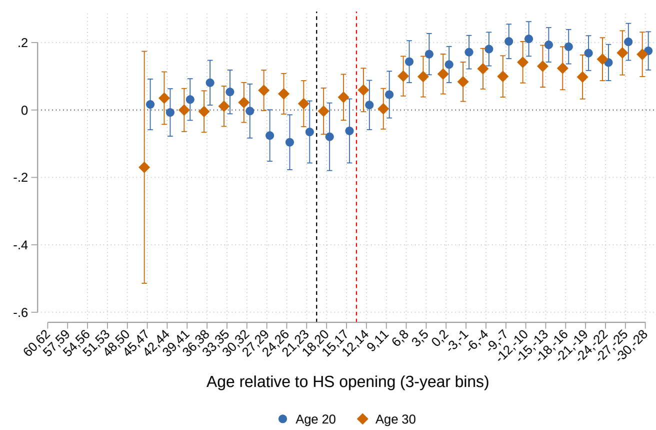
(b) Without county-by-year FEs



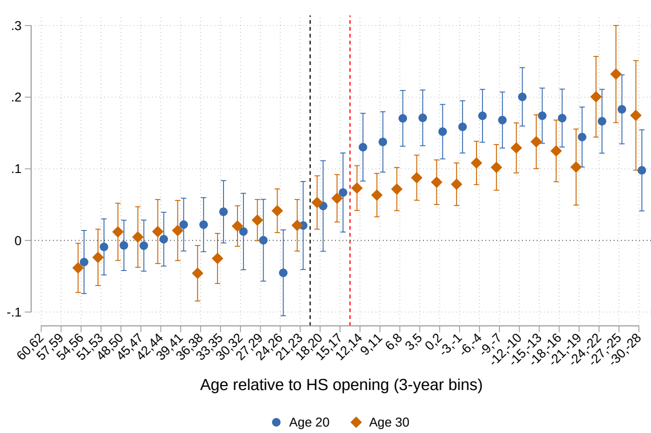
(c) Include HS-by-year FEs



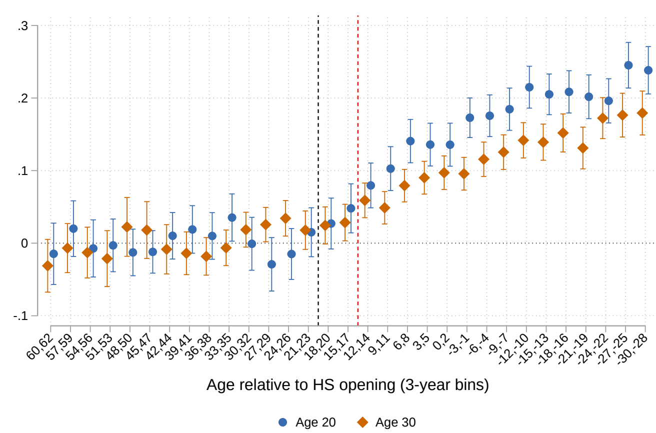
(d) Non-movers



(e) More balanced city panel



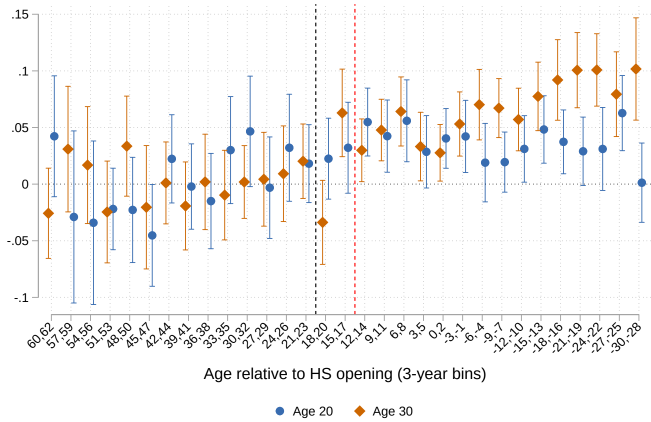
(f) With parental occupation and RR controls



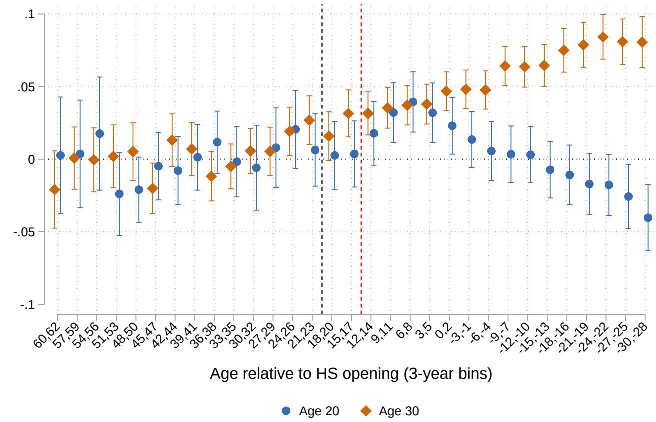
Notes: This figure shows our event study results for each indicated robustness check the indicated age groups. See Equation 2 for specification and sample details.

Figure A31: Log occupational income robustness, men

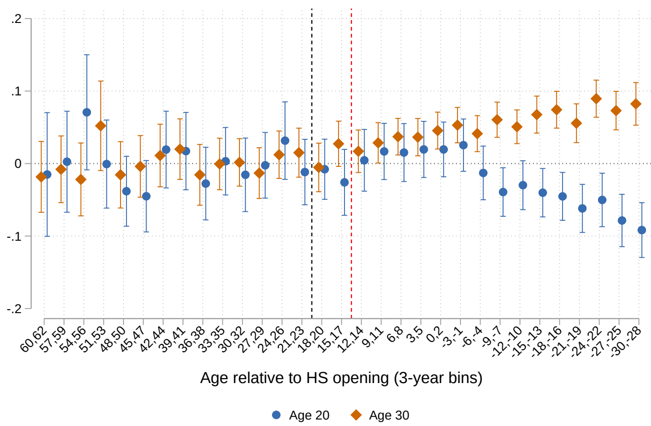
(a) Only link 17-18 year olds



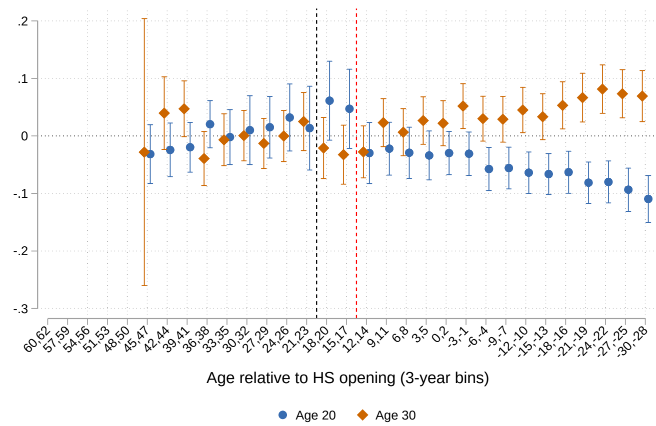
(b) Without county-by-year FEs



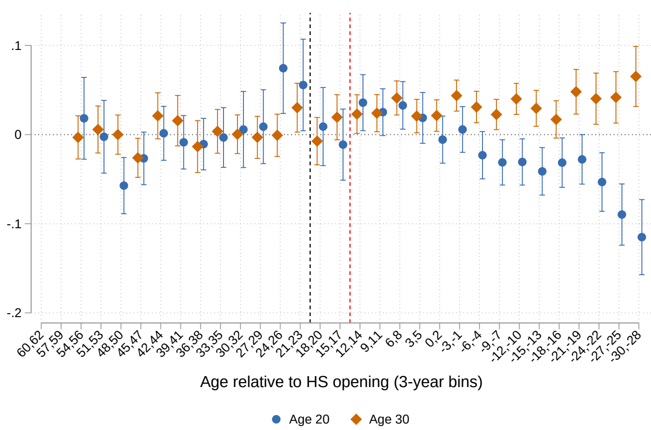
(c) Include HS-by-year FEs



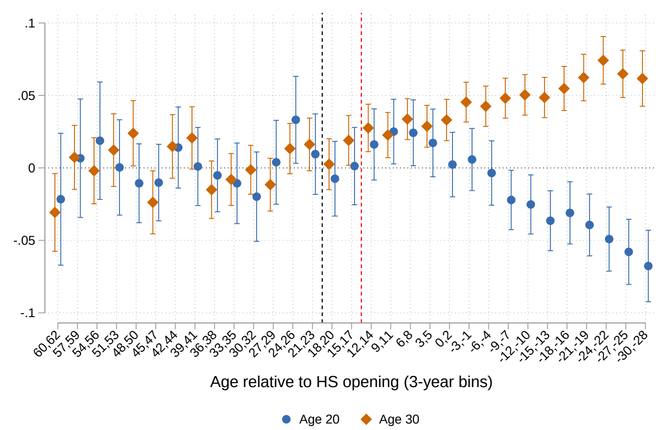
(d) Non-movers



(e) More balanced city panel



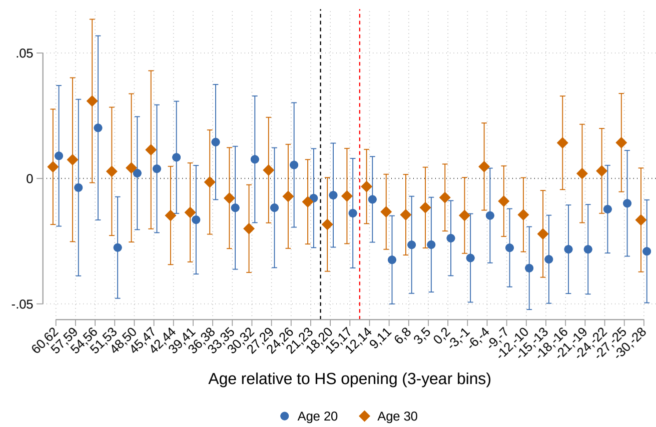
(f) With parental occupation and RR controls



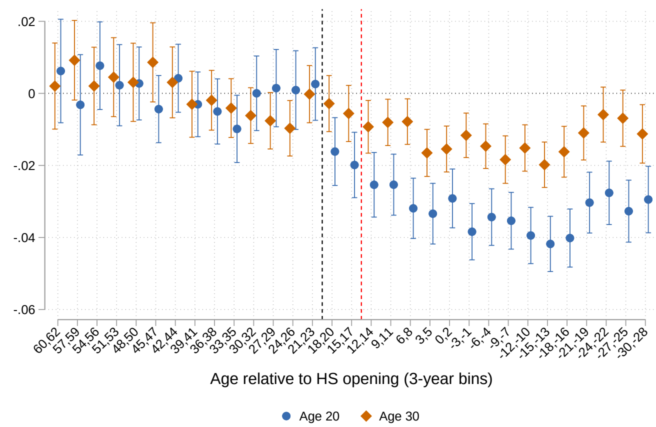
Notes: This figure shows our event study results for each indicated robustness check the indicated age groups. See Equation 2 for specification and sample details.

Figure A32: Marriage robustness, women

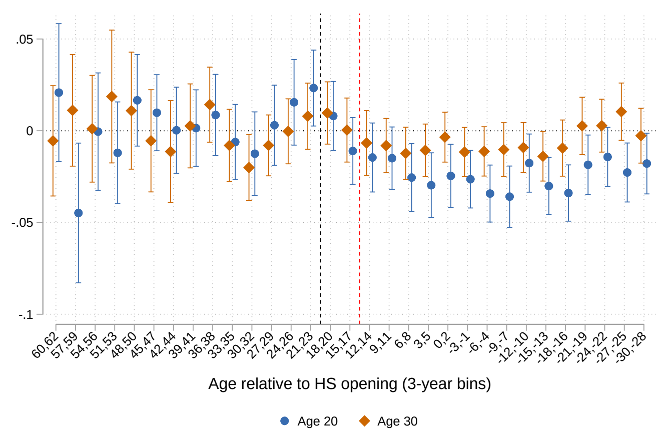
(a) Only link 17-18 year olds



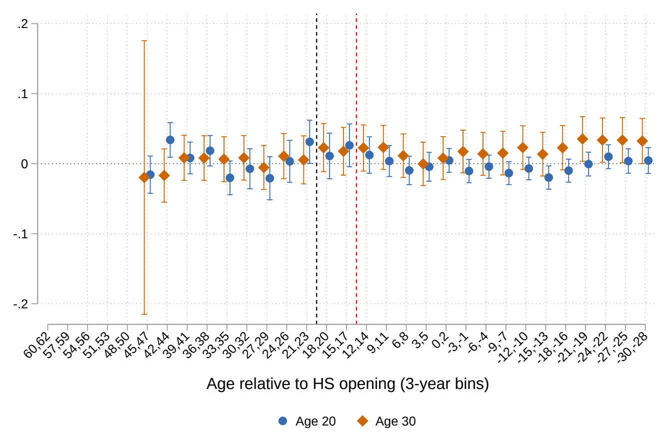
(b) Without county-by-year FEs



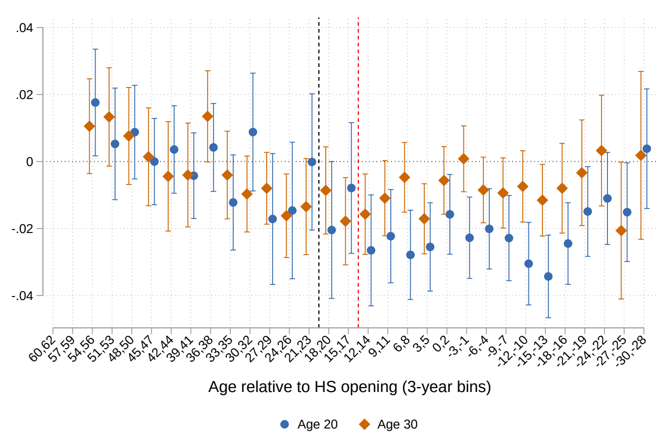
(c) Include HS-by-year FEs



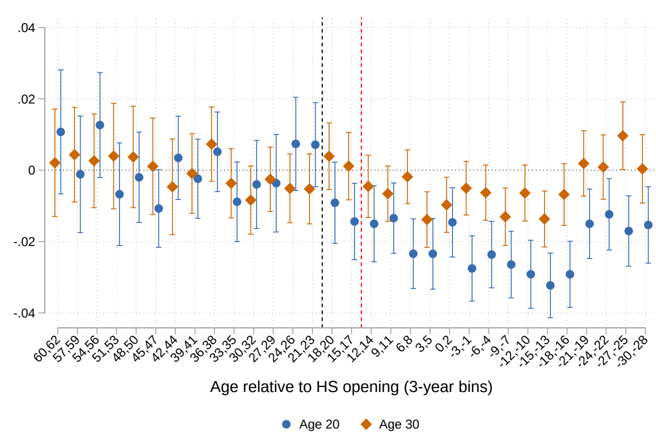
(d) Non-movers



(e) More balanced city panel



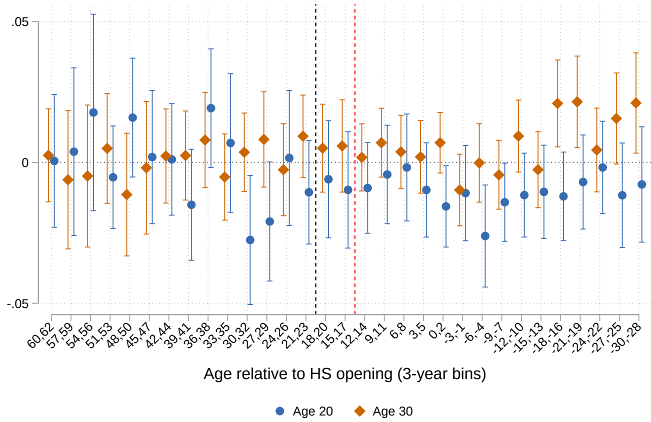
(f) With parental occupation and RR controls



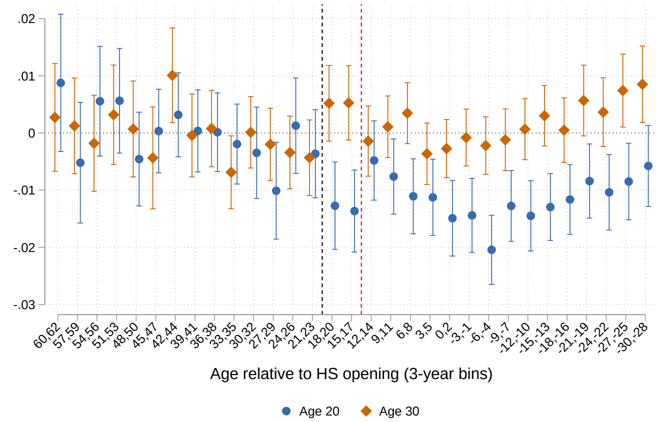
Notes: This figure shows our event study results for each indicated robustness check the indicated age groups. See Equation 2 for specification and sample details.

Figure A33: Marriage robustness, men

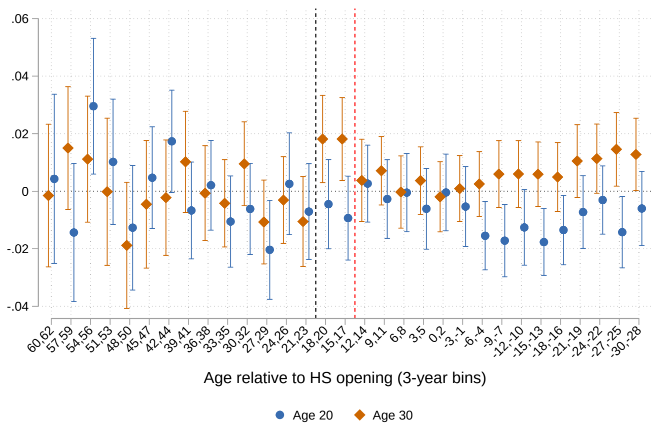
(a) Only link 17-18 year olds



(b) Without county-by-year FEs

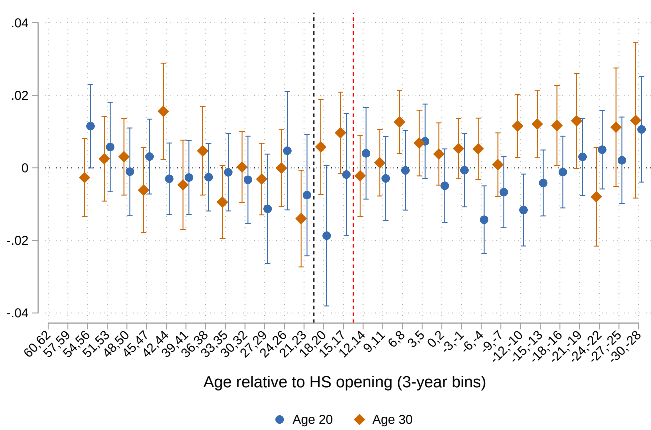


(c) Include HS-by-year FEs

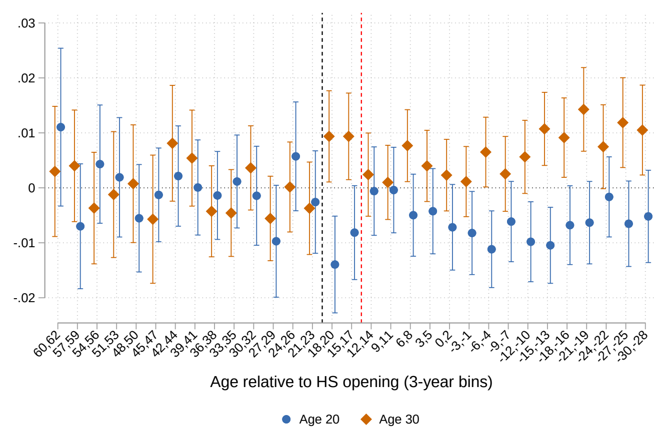


(d) Non-movers

(e) More balanced city panel

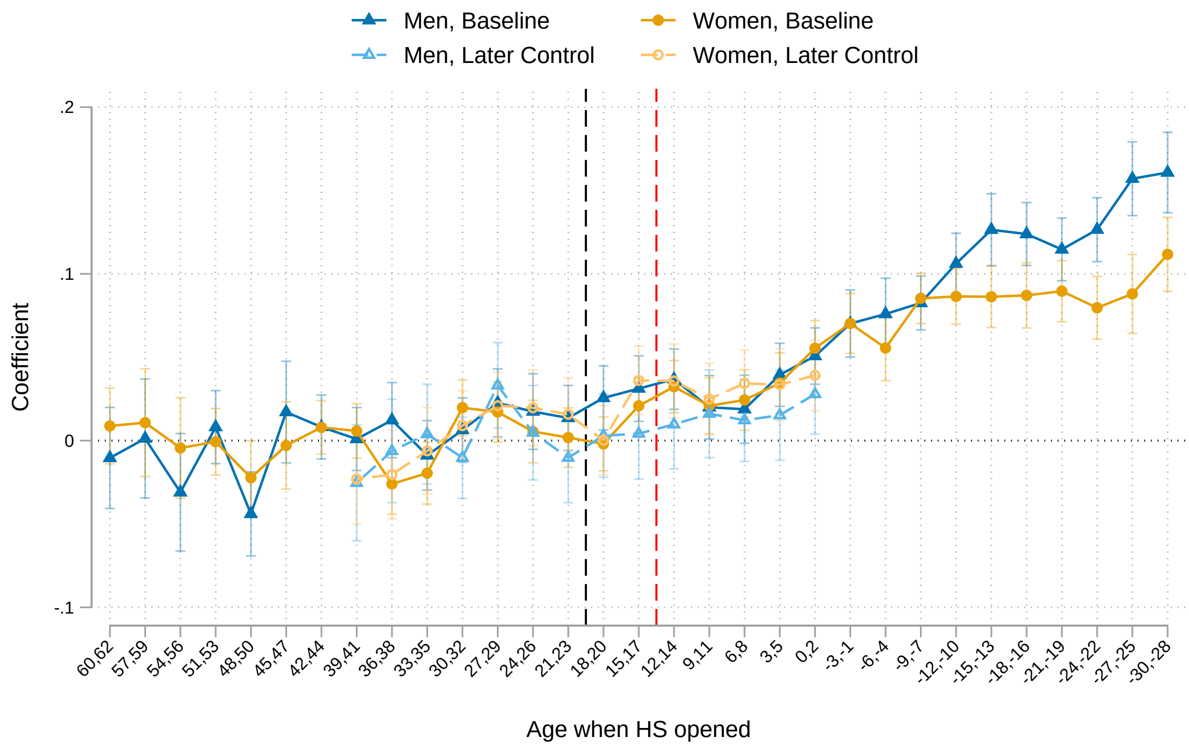


(f) With parental occupation and RR controls



Notes: This figure shows our event study results for each indicated robustness check the indicated age groups. See Equation 2 for specification and sample details.

Figure A34: Later-treated high school cities as control for earlier-treated high school cities, enrollment

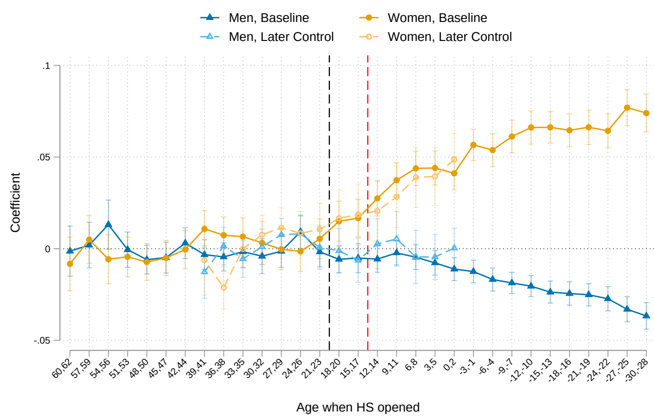


Notes: This figure shows our baseline enrollment result overlaid with an alternative specification that uses places with later-treated high schools as a control for places with earlier-treated high schools. Results are shown for both men and women. We assign later-treated high school cohort cities (e.g., places that built a high school in 1900) as controls for cities that built high schools 21 years earlier (e.g, places that built a high school in 1879). In this case, we define 1871 as the focal cohort year. We then assign treatment timing relative to the focal cohort year for both the treatment and control group and estimate versions of Equation 1 that condition on place-by-focal cohort, year-by-focal cohort, and birth-year-by-focal cohort fixed effects. Standard errors clustered by the nearest high school.

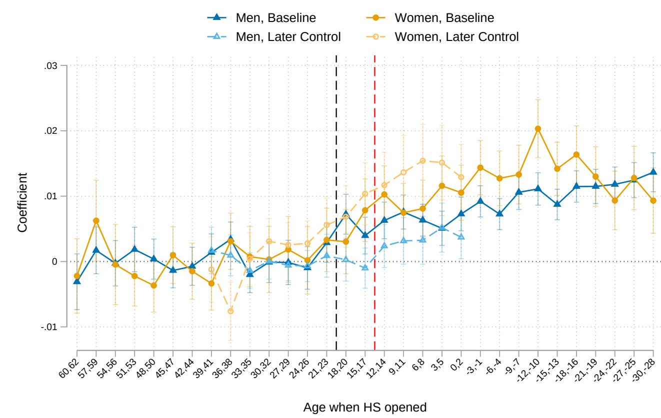
The estimation window for this analysis is shorter than the baseline because we only use early-treated units in the treatment group (since we need the later treated units as controls) and because the post-period needs to be truncated so that we avoid picking up a treatment effect for the control units as they build their own high schools.

Figure A35: Later-treated high school cities as control for earlier-treated high school cities, age 20s results

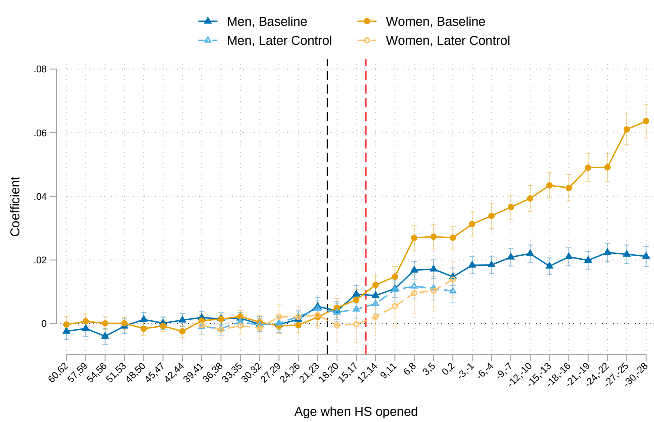
(a) Labor force participation



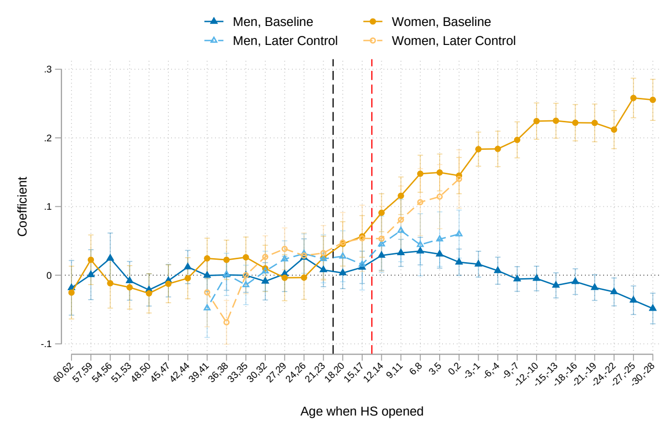
(b) Professional



(c) Clerk



(d) Log occupational score

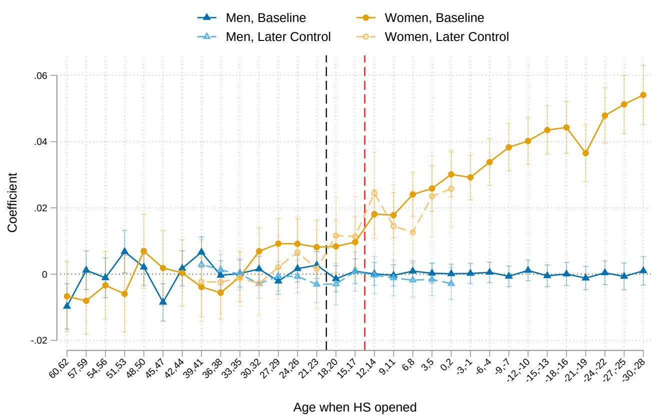


Notes: These panels show our baseline result for the indicated age 20s outcome overlaid with an alternative specification that uses places with later-treated high schools as a control for places with earlier-treated high schools. Results are shown for both men and women. We assign later-treated high school cohort cities (e.g., places that built a high school in 1900) as controls for cities that built high schools 21 years earlier (e.g, places that built a high school in 1879). In this case, we define 1871 as the focal cohort year. We then assign treatment timing relative to the focal cohort year for both the treatment and control group and estimate versions of Equation 2 that condition on place-by-focal cohort, year-by-focal cohort, and birth-year-by-focal cohort fixed effects. Standard errors clustered by the nearest high school.

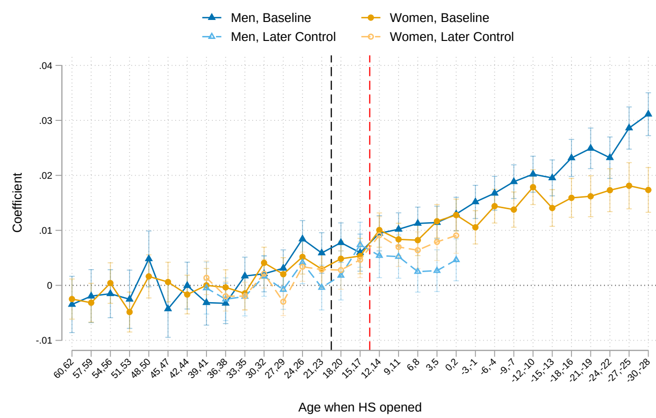
The estimation window for this analysis is shorter than the baseline because we only use early-treated units in the treatment group (since we need the later treated units as controls) and because the post-period needs to be truncated so that we avoid picking up a treatment effect for the control units as they build their own high schools.

Figure A36: Later-treated high school cities as control for earlier-treated high school cities, age 30s results

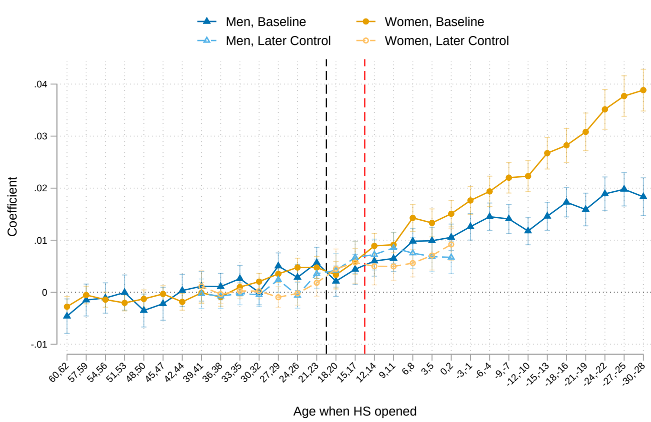
(a) Labor force participation



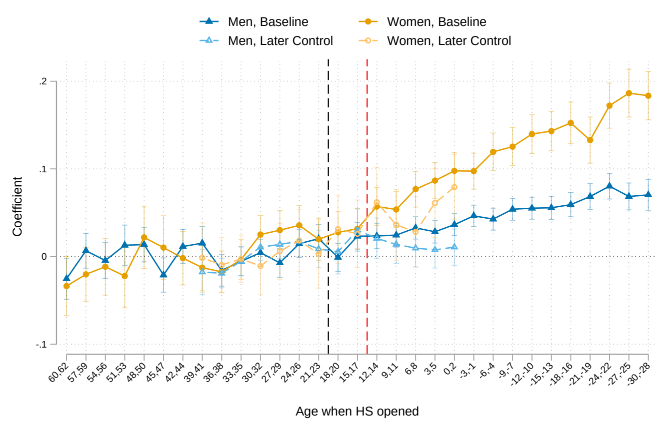
(b) Professional



(c) Clerk



(d) Log occupational score

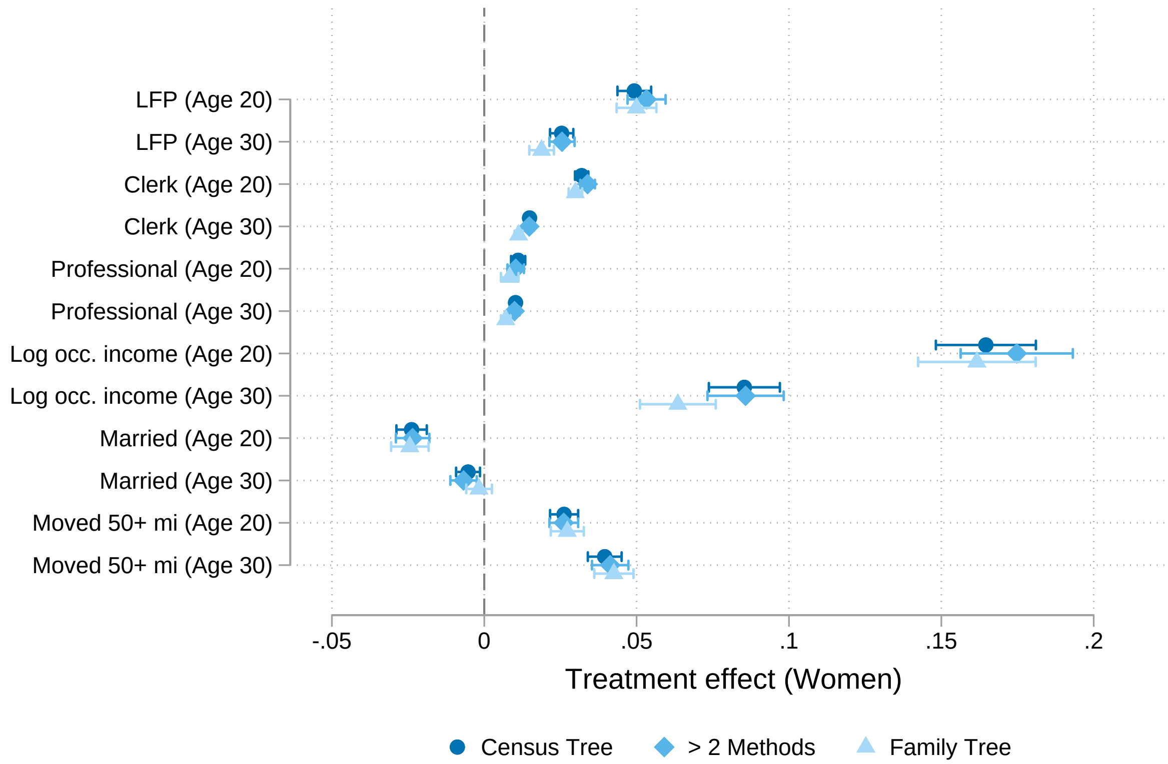


Notes: These panels show our baseline result for the indicated age 30s outcome overlaid with an alternative specification that uses places with later-treated high schools as a control for places with earlier-treated high schools. Results are shown for both men and women. We assign later-treated high school cohort cities (e.g., places that built a high school in 1900) as controls for cities that built high schools 21 years earlier (e.g, places that built a high school in 1879). In this case, we define 1871 as the focal cohort year. We then assign treatment timing relative to the focal cohort year for both the treatment and control group and estimate versions of Equation 2 that condition on place-by-focal cohort, year-by-focal cohort, and birth-year-by-focal cohort fixed effects. Standard errors clustered by the nearest high school.

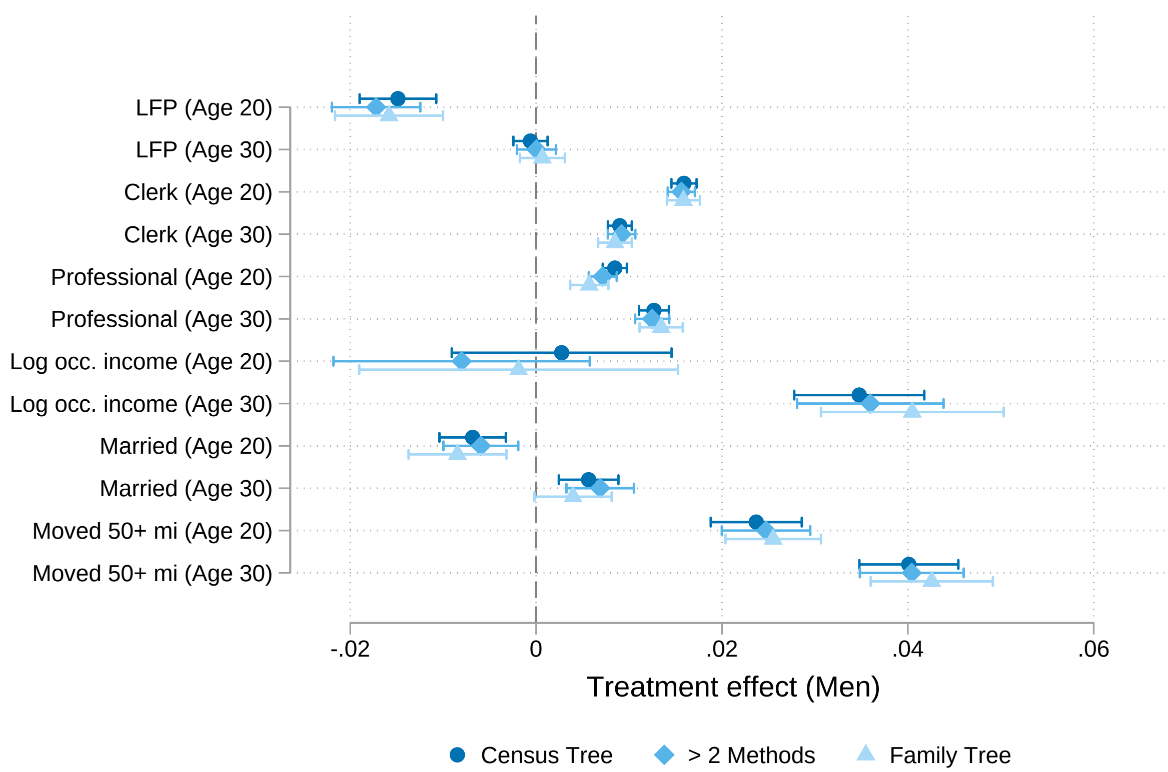
The estimation window for this analysis is shorter than the baseline because we only use early-treated units in the treatment group (since we need the later treated units as controls) and because the post-period needs to be truncated so that we avoid picking up a treatment effect for the control units as they build their own high schools.

Figure A37: Linking method robustness, by gender and long run outcome

(a) Women

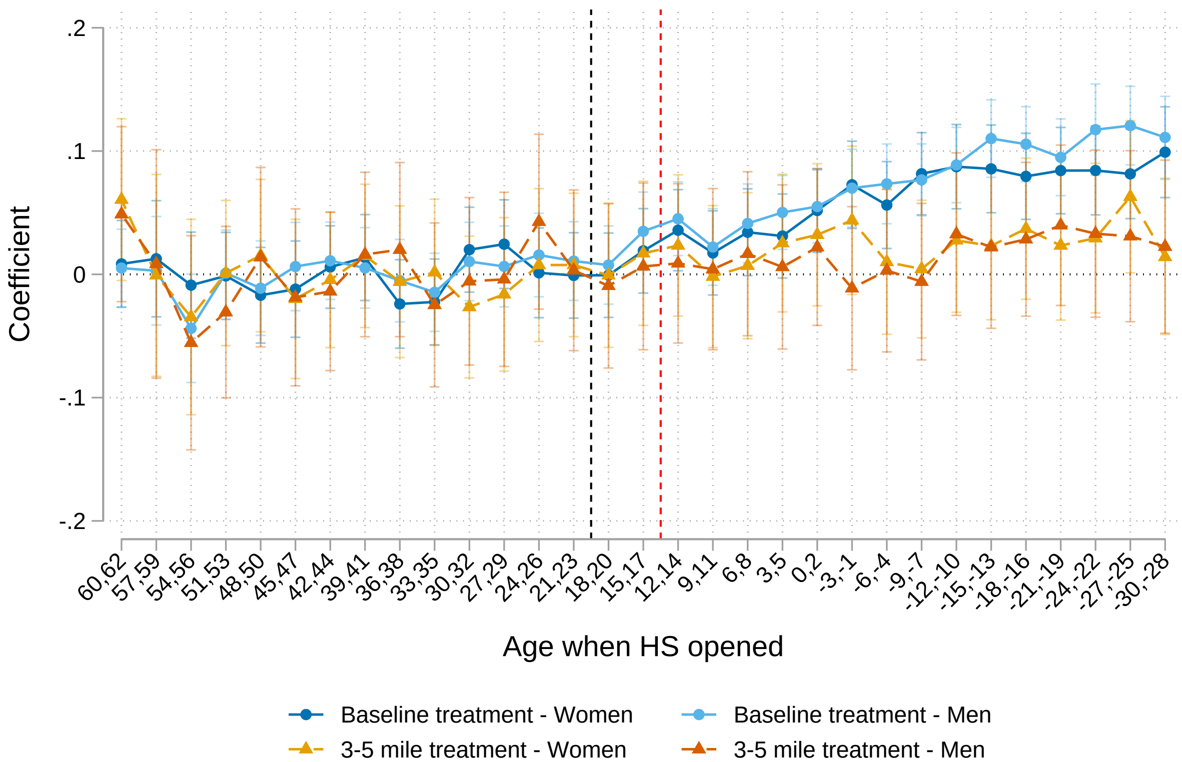


(b) Men



Notes: See Figure 6 notes for specification details. "> 2 methods" indicates that the link was made using more than one method in the Census Tree. "Family Tree" uses just genealogical links.

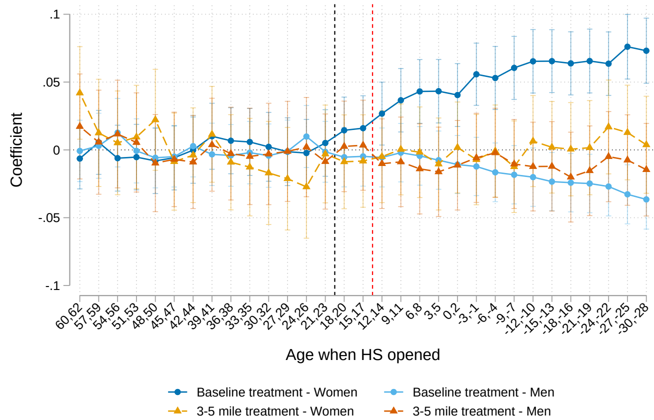
Figure A38: 3-5 mile treatment definition compared to baseline 0.5 mile treatment, attendance



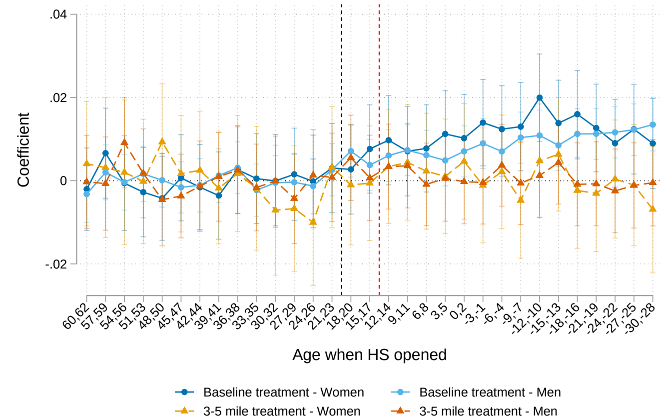
Notes: This figure shows our baseline attendance result compared to an alternative treatment group: those who are within 3-5 miles of the high school. See Figure 2 notes for additional specification details.

Figure A39: 3-5 mile treatment definition compared to baseline 0.5 mile treatment, age 20 outcomes

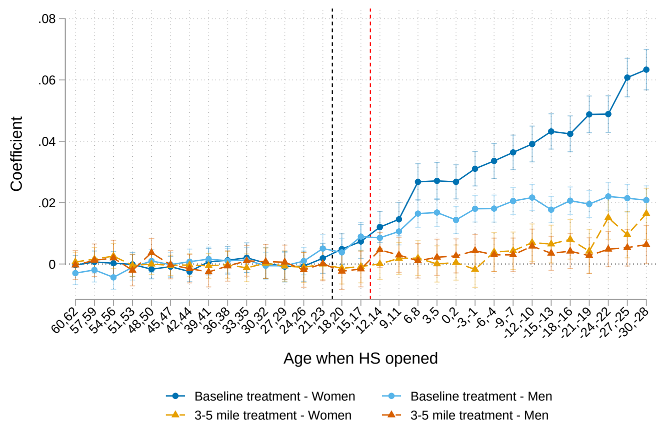
(a) Labor force participation



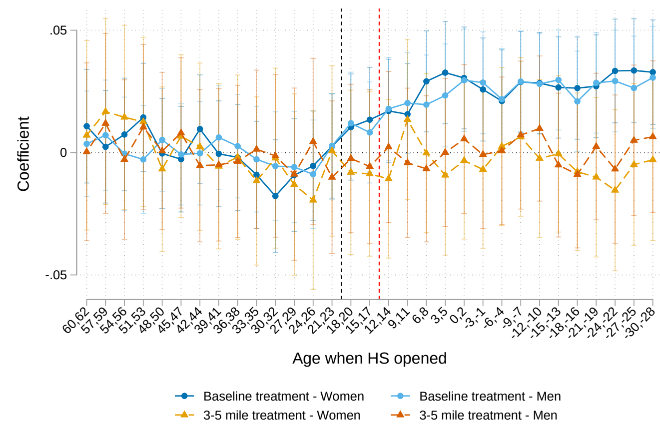
(b) Professional



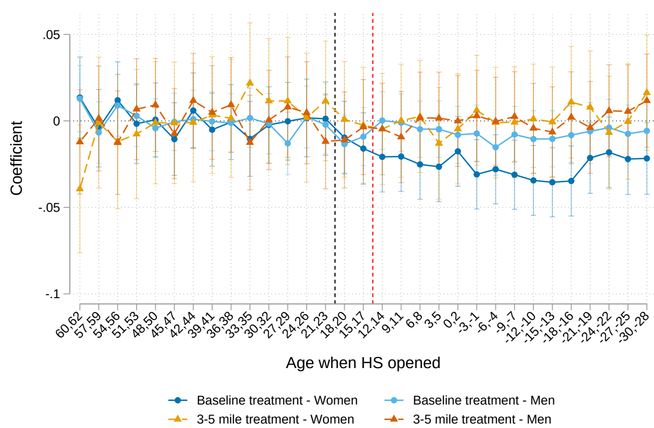
(c) Clerk



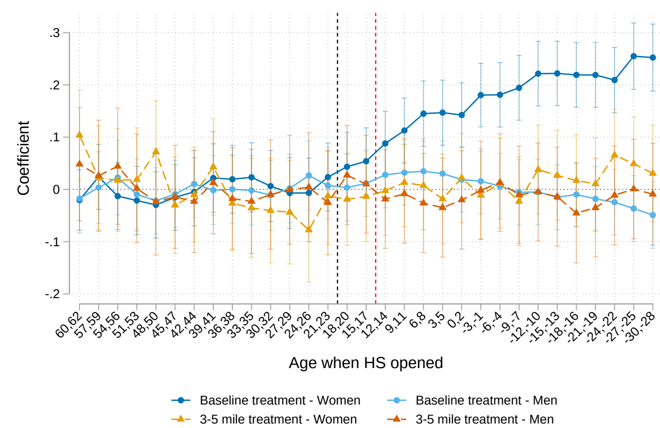
(d) Geographic mobility



(e) Marriage

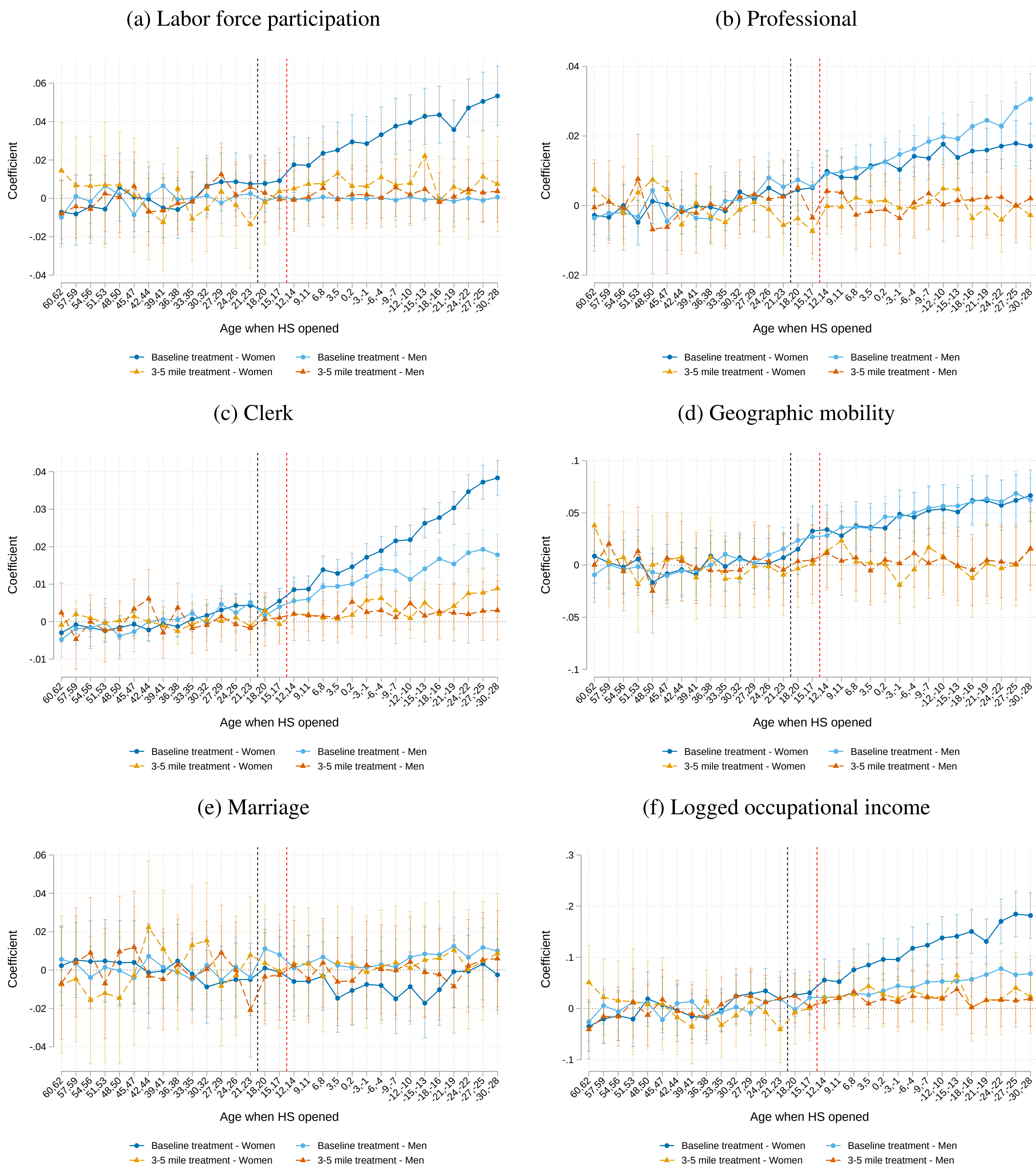


(f) Logged occupational income



Notes: This figure shows our baseline long run results compared to an alternative treatment group: those who are within 3-5 miles of the high school. See Figure 4 notes for specification details.

Figure A40: 3-5 mile treatment definition compared to baseline 0.5 mile treatment, age 30 outcomes



Notes: This figure shows our baseline long run results compared to an alternative treatment group: those who are within 3-5 miles of the high school. See Figure 4 notes for specification details.