When Less Still Harmed: Old-Age Mortality Effects of In Utero Exposure to the 1918 Influenza Pandemic for Males in Low-Intensity Cities[1]

By MUZHE YANG[2]

March 2026

Preliminary Draft: Please do not cite or quote without the author's permission.

Abstract

Using city-level excess influenza mortality to proxy pandemic intensity, this study examines the impact of in utero exposure to the 1918 influenza pandemic on male old-age mortality. Census-linked death records, merged to the 1920 full-count census and linked censuses through 1950, enable a within-city adjacent-cohort comparison of 1919 (treated) versus 1918 (control) births, conditioning on detailed background characteristics and city and birth month fixed effects. Estimates suggest about a three-month longevity deficit in low-intensity cities and null effects in high-intensity cities, consistent with scarring versus selection. Mechanism evidence points to poorer entry jobs and persistently lower wages.

Keywords: pandemic, mortality, longevity, 1918 influenza, scarring effects
JEL classification: I10, I14, I15, I18, J10, J14, J24, C21, N32, N92

[1] The data used in this study were drawn from the Social Security Administration mortality records linked to the full-count 1940 U.S. Census through CenSoc, available at https://censoc.berkeley.edu/data/ (accessed in September 2025), and from U.S. Census data provided by IPUMS, available at https://usa.ipums.org/usa/ (accessed in September 2025). Under the data use agreements of both CenSoc and IPUMS, the author is not permitted to redistribute the downloaded data files. The paper provides the relevant source information and citations for these datasets to allow replication.

[2] Corresponding Author, Department of Economics, College of Business, Lehigh University; 621 Taylor Street, Bethlehem, PA 18015; muzheyang@lehigh.edu. Declarations of relevant or material financial interests that relate to the research described in this paper: none. Declarations of conflicts of interest: none.

I. Introduction

The COVID-19 pandemic has renewed interest in the 1918 influenza pandemic, "the most recent historical event to match or exceed the scale of the COVID-19 pandemic" (Beach, Clay and Saavedra 2022, 43). Because it occurred more than a century ago, the 1918 pandemic provides a rare opportunity to study life-course impacts of early-life shocks, including in utero exposure (e.g., Beach et al. 2022). Existing work has linked prenatal exposure to later human capital and socioeconomic outcomes, such as educational attainment and adult earnings (Almond 2006), as well as health outcomes from childhood through older ages, including height and later-life health problems (Lin and Liu 2014). This study extends that literature by focusing on old-age[1] mortality (longevity). Longevity is a particularly informative outcome, as it captures the ultimate long-run health consequence, and it is a key indicator of social welfare (Lutz et al. 2021).

With this life-course perspective in mind, a key challenge is credible identification of the pandemic's long-run effects. Almond's (2006) seminal work treats the fall 1918 wave of the pandemic as a sharp and short-lived shock and compares the 1919 birth cohort with adjacent cohorts to limit secular trend confounding. While this comparison captures meaningful variation in exposure, it may also pick up cross-cohort differences driven by other contemporaneous shocks, most notably World War I and related disruptions. In that context, a simple comparison of adjacent birth cohorts may not necessarily isolate exogenous variation in pandemic exposure. Beach et al. (2022) highlight this concern by showing that the 1919 birth cohort is negatively selected on parental socioeconomic status, and the estimated 1919 deficits in later outcomes (e.g., years of education) attenuate substantially and become statistically insignificant once background characteristics are controlled for.

Even with extensive controls, residual confounding may remain if the pandemic coincided with other shocks that are imperfectly observed. A way forward, emphasized in both Almond (2006) and Beach et al. (2022), is to exploit substantial geographic variation in pandemic intensity using measures that isolate unexpected local severity. In this regard, Beach et al. (2022) provide evidence that city-level excess influenza mortality is plausibly exogenous and construct a measure that

[1] In this study we use 65 as the age cutoff for old age, following the guidance by the U.S. National Institute on Aging (https://www.nih.gov/nih-style-guide/age, accessed in January 2026). This threshold is particularly salient in the United States because Medicare is health insurance for people aged 65 or older, making 65 a practical marker for entry into the elderly population (https://www.medicare.gov/basics/get-started-with-medicare, accessed in January 2026).

captures the unexpected 1918 spike relative to each city's pre-1918 trend. Building on this work, this study addresses selection into birth cohorts by conditioning on an extensive set of parental and household background characteristics and also by utilizing plausibly exogenous city-level variation in pandemic intensity.

This study examines whether in utero exposure to the 1918 influenza pandemic reduces old-age longevity, even where the pandemic's intensity was low. Focusing on low intensity does not mean assuming small shock. Low intensity can coincide with meaningful fetal insults[2] if high-intensity places induce stronger private protective behaviors (e.g., avoidance and social distancing) that reduce the effect of marginal exposure, whereas low-intensity places may not, so marginal exposure could translate more directly into fetal damage. Furthermore, low intensity may result from stronger institutional responses, that is, non-pharmaceutical interventions such as quarantine and social distancing used to control the spread (Bootsma and Ferguson 2007; Correia, Luck and Verner 2022; Hatchett, Mecher and Lipsitch 2007), which may themselves elevate stress and disruption during pregnancy. Consistent with a scarring versus selection (culling)[3] framework, high-intensity cities may experience stronger in utero and infant mortality selection (culling), so survivors are positively selected on biological resilience (e.g., stronger immune systems), muting detectable long-run mortality impacts.[4] In contrast, in low-intensity cities the pandemic shock may not be strong enough to induce as much fetal loss or infant mortality as in high-intensity cities, leaving a survivor cohort that include more individuals who are alive but scarred and whose later-life outcomes (e.g., longevity) may be compromised. Importantly, fetal insults need not be limited to direct maternal infection; they may also operate through other channels such as maternal stress,[5] which can matter even where pandemic intensity was modest. It is this low-intensity setting that has received limited attention in prior work.

---

[2] Fetal insults refer to adverse events or conditions that negatively affect the developing fetus during pregnancy, such as maternal infections, exposure to pollutants, and stress (Almond and Currie 2011).

[3] Scarring refers to lasting damage of fetal insults among survivors, whereas selection (culling) occurs when a severe shock increases fetal loss, infant mortality, or early-life mortality, leaving a surviving cohort that has better baseline health.

[4] Unlike the historically U-shaped age-mortality pattern for influenza (i.e., high mortality among the very young and the very old), the 1918 pandemic showed a distinctive W-shaped pattern, with an additional mortality peak among young adults aged 20–40 (source: https://wwwnc.cdc.gov/eid/article/12/1/05-0979_article, accessed in January 2026). This unusual concentration of deaths among prime-age adults implies that selective mortality was not limited to fetuses and infants; some frailer individuals may have been culled even in young adulthood, leaving survivors who reach older ages as a more positively selected cohort.

[5] Aizer, Stroud and Buka (2016) show that maternal stress during pregnancy can negatively affect a wide range of offspring outcomes including cognition, health, and educational attainment.

Empirically, this study uses the city-level excess influenza mortality measure constructed by Beach et al. (2022) to classify cities into low- and high-intensity groups and combines within-city adjacent-cohort comparisons with this cross-city intensity variation. We merge the 1940 census linked to the Social Security Administration (SSA) Death Master File (DMF), commonly referred to as CenSoc-DMF[6] (Goldstein et al. 2025), with the 1920 full-count census, the 1920–1940 linked census, and the 1940–1950 linked census. A key advantage of CenSoc-DMF (version 4.0, the most recent version at the time of this study) is that it provides exact birth year and month and covers deaths from 1975 to 2020, reducing truncation for the 1918–1919 cohorts and enabling clean adjacent-cohort comparisons and using birth month fixed effects to control for seasonality in birth outcomes (Currie and Schwandt 2013). Furthermore, our assembled dataset serves three purposes. First, it provides a rich set of background characteristics, mitigating concerns that cohorts exposed in utero to the pandemic are selected with respect to parental characteristics. Second, using the exact birth year information in CenSoc-DMF, this study extends prior research in which birth year was inferred from census data by comparing adjacent cohorts: the 1919 birth cohort (treated) and the 1918 birth cohort (control). Conditional on city-level excess influenza deaths, background characteristics appear balanced between the treated and control groups, and the estimated treatment effects are robust to controlling for background covariates, city and birth month fixed effects. Third, the linked census data allow individuals to be tracked over decades, enabling a direct examination of long-run mechanisms by following individuals from childhood circumstances (1920) through early-adult labor-market outcomes (1940) and midlife outcomes (1950).

Our findings align with a scarring versus selection interpretation. In low-intensity cities, males born in 1919 live about three months less than those born in 1918—an effect comparable in magnitude to the long-run longevity impact of losing one year of schooling documented in prior work (Halpern-Manners et al. 2020). In high-intensity cities, estimated mortality effects are smaller and statistically indistinguishable from zero, consistent with selection offsetting scarring and making net effects harder to detect. Falsification tests show no effects in other adjacent-cohort comparisons. The findings are robust to Lee bounds (Lee 2009) taking into account sample selection, Oster's sensitivity analysis (Oster 2019) accounting for selection on unobservables, and

[6] It is called CenSoc-DMF because the CenSoc project (short for linking the 1940 Census "Cen" to Social Security mortality records "Soc") created a version of its linked mortality data where the death information comes specifically from the Social Security DMF (Breen, Osborne and Goldstein 2023).

alternative definitions of high- and low-intensity areas. Mechanism analyses suggest that reduced longevity could operate in part through labor-market channels: exposed cohorts enter adulthood in lower-paid jobs and remain on persistently lower-wage trajectories.

Methodologically, this paper also shows how causal machine learning (ML) tools, specifically the generalized random forest (Athey, Tibshirani and Wager 2019) and double/debiased ML (Chernozhukov et al. 2018; Kennedy 2023; Knaus 2022; Semenova and Chernozhukov 2021), can uncover treatment effect heterogeneity in settings where conventional subgroup analyses are under-powered because of small group size. Results indicate that parental education could be an important moderator of the longevity impact.

Finally, this study's focus on males is a limitation driven by the linkage structure of CenSoc-DMF and other linked census data, but this focus is also substantively relevant. Consistent with the fragile male hypothesis (Eriksson et al. 2010; Kraemer 2000), evidence suggests that severe prenatal pandemic exposure may reduce the proportion of male births (Ogasawara 2022), that is, male fetal loss under intense fetal insults, implying that under low intensity, where there can be less culling, more damaged but viable males may survive to adulthood and carry scarring into old age. A key policy implication is that low intensity does not guarantee low fetal risk: in low-intensity areas, there still can be elevated stress which, through in utero exposure, can have long-term adverse effects. It is important to have public-health policies protecting and supporting pregnant women even when local pandemic severity appears modest.

The rest of the paper proceeds as follows. Section 2 reviews the literature and provides a brief overview of the 1918 influenza pandemic in the United States, highlighting features that motivate our identification strategy, which we discuss in detail in Section 3. Section 4 presents the main results, robustness checks, mechanisms and heterogeneity analyses. Section 5 concludes.

II. Background

A. Overview of the 1918 Influenza Pandemic in the United States

In this section we highlight three defining features of the 1918 influenza pandemic in the United States that motive our identification strategy (which we explain in detail in Section 3). First, the pandemic unfolded in multiple waves: a relatively mild spring–summer 1918 wave (i.e., the first wave) was followed by a highly lethal fall 1918 wave (i.e., the second wave), which peaked

between September and November and accounted for most U.S. pandemic deaths; a further wave (i.e., the third wave) occurred in winter 2018–spring 2019 and subsided by summer 2019.[7] This timing motivates comparing adjacent birth cohorts (1919 vs. 1918). Because full-term gestation lasts about 40 weeks, approximately nine months, many 1919 births, especially early 1919 births (or mid to late 1919 births), were in utero during the fall 1918 peak (or the winter 2018–spring 2019 resurgence); in contrast, 1918 births were more likely to have completed gestation before the fall 1918 peak and thus had less prenatal overlap with the pandemic's most lethal months.

Second, the pandemic's severity varied sharply across locations rather than being uniform nationally (Beach, Clay and Saavedra 2022; Bootsma and Ferguson 2007; Hatchett, Mecher and Lipsitch 2007; Markel et al. 2007). Spread was uneven and mortality differed substantially across cities, in part because cities responded differently, for example in the timing and duration of non-pharmaceutical interventions (Beach, Clay and Saavedra 2022; Bootsma and Ferguson 2007; Hatchett, Mecher and Lipsitch 2007). Importantly, these policy choices may not track a city's long-run economic trajectory (see also Correia, Luck and Verner 2022); instead, they may reflect local decision-making under specific circumstances. Markel et al. (2007) illustrate this contrast with Philadelphia and New York City: Philadelphia acted late, introducing interventions only after mortality had already accelerated, whereas New York City acted early and sustained measures such as isolation and quarantine, achieving "the lowest death rate on the Eastern seaboard" (p. 651). This kind of plausibly idiosyncratic policy variation, potentially translating into unexpected differences in excess influenza deaths, motivates using cross-city variation in pandemic intensity, rather than relying solely on birth timing.

Third, the pandemic coincided with U.S. participation in World War I, raising concerns about selection and confounding in simple adjacent-cohort comparisons. Large-scale mobilization and overseas deployment during summer 1918 through early 1919 overlapped in time with the pandemic.[8] Beach et al. (2022) emphasize that wartime conditions can affect the composition of births, for example, through selection into fatherhood (since males who served in the military were positively selected on socioeconomic status) and through fertility responses; therefore, adjacent cohorts may differ in family background.

---

[7] Sources: https://archive.cdc.gov/www_cdc_gov/flu/pandemic-resources/1918-commemoration/pandemic-timeline-1918.htm (accessed in January 2026).

[8] Source: https://archive.cdc.gov/www_cdc_gov/flu/pandemic-resources/1918-commemoration/pandemic-timeline-1918.htm (accessed in January 2026).

Taken together, these historical features motivate our study's identification strategy for the longevity effect of in utero exposure to the 1918 influenza pandemic, which is combining within-city adjacent-cohort comparisons (1919 vs. 1918) with plausibly exogenous cross-city variation in unexpected local severity, while controlling for an extensive set of parental and household characteristics and fixed effects.

B. Literature Review

One important motivation for examining in utero exposure to the 1918 influenza pandemic is that it provides a historically large, plausibly unexpected health shock with sharply defined timing. This setting has been used to test the fetal origins hypothesis; that is, in utero shocks can permanently alter later-life health and disease risks (Almond and Currie 2011; Barker 1995). A seminal work by Almond (2006) emphasizes that the pandemic's rapid onset and short duration generate sharp predictions for those in utero during the peak months; using 1960–1980 census data, Almond (2006) finds that the exposed cohort has lower educational attainment, higher rates of physical disability, lower income and socioeconomic status in adulthood. Studies following Almond (2006) expand the outcome set to include later-life health and mortality, and they emphasize that identification of the pandemic's impact can be sensitive to design choices and cohort compositions. For example, Mazumder et al. (2010) show that individuals whose birth timing corresponds to peak prenatal exposure to the 1918 pandemic exhibit increased risks of later-life cardiovascular diseases; using both annual and quarter-of-birth variation to identify in utero exposure, they also find small but statistically significant reductions in male adult height in World War II enlistment records for the 1919 birth cohort. Myrskylä, Mehta and Chang (2013) examine old-age mortality by cause of death and find that individuals born during the peak of the 1918 pandemic experience excess all-cause mortality at old ages, consistent with long-term biological or developmental scarring from early-life exposure; they also point out that those who survive into adulthood may be positively selected, complicating the interpretation of later-life mortality patterns. Furthermore, Fletcher (2018) shows no consistent evidence that specific gestational windows of exposure matter. In contrast, Helgertz and Bengtsson (2019) find that for socioeconomic outcomes (e.g., income and occupation), first-trimester exposure can be positively associated, likely reflecting selection effects rather than any direct benefit of exposure; for hospitalization risk, second-trimester exposure may be consequential for males, while third-

trimester exposure may matter for females. Additional work using non-U.S. settings also finds adverse impacts on later outcomes. For example, exposed cohorts in Switzerland show lower educational attainment and a lower likelihood of marriage (Neelsen and Stratmann 2012), while in Brazil they are less likely to be college educated and employed and earn lower hourly wages (Nelson 2010). Taken together, these findings suggest that fetal-origin effects of the 1918 pandemic can arise across distinct institutional contexts.

In this literature, an important conceptual distinction is between scarring and selection (culling). Scarring refers to lasting damage of fetal insults among survivors, whereas selection (culling) occurs when a severe shock increases fetal loss, infant mortality, or early-life mortality, leaving a surviving cohort that is positively selected on underlying health and resilience. This tension is emphasized in economics (e.g., Almond and Currie 2011) and demography (e.g., Bozzoli, Deaton and Quintana-Domeque 2009) as a general challenge for interpreting the long-run impacts of severe early-life shocks. In the specific context of the 1918 pandemic, many studies discuss selection but do not focus on an explicit comparison between low- and high-intensity places. Instead, intensity is often considered either implicitly, because cohort timing is assumed to proxy exposure to the pandemic's health shock, or treated as a continuous variable through dose-response designs that relate outcomes to local (or regional) mortality-based proxies for exposure. For example, Almond (2006) combines cross-cohort comparisons with a dose-response approach using mortality-based proxies for exposure intensity. Beach et al. (2022) reevaluate Almond's (2006) cross-cohort and dose-response evidence, and they add new evidence using city-level intensity variation; however, this intensity variation is mainly used to strengthen identification and check robustness rather than to treat the low-intensity case as a substantively distinct setting.

The scarring versus selection distinction is also central to our male-focused analysis. Although the restriction to males is directly driven by data limitations, it is substantively relevant in light of the fragile male hypothesis (Eriksson et al. 2010; Kraemer 2000), which posits that males are biologically more vulnerable from conception onward, more susceptible in utero to adverse maternal environments and stressors, and exhibit higher mortality throughout the life course. In the context of prenatal pandemic exposure, this vulnerability implies that cohort composition may vary with shock intensity: harsher in utero conditions can selectively cull the most vulnerable male fetuses, altering the surviving cohort even before birth. Related evidence links pandemic exposure to shifts in the sex ratio at birth. Using data from the 1918–1920 influenza pandemic in Japan,

Ogasawara (2022) finds that fetal exposure during the first trimester reduces the proportion of male births, consistent with male-biased fetal loss under severe in utero insults. This reinforces a key point for long-run outcome studies: estimated effects among observed survivors may reflect both scarring and sex-specific selection, and those forces may not move in the same direction. In related U.S. evidence on later-life health, Mazumder et al. (2010) also report patterns consistent with sex differences in later-life disease responses to pandemic-era birth timing. Specifically, Mazumder et al. (2010) find that males born in the first quarter of 1919, who were likely exposed to the 1918 influenza pandemic during the second trimester of pregnancy, show significantly higher prevalence of total heart disease (23.1%) and ischemic heart disease (32.7%) compared with females. This aligns with the idea that males may be a particularly informative margin for detecting persistent consequences of early-life health shocks.

Empirical strategies used to identify the impact of in utero exposure to the 1918 pandemic largely fall into two categories: cross-cohort timing comparisons and dose-response designs exploiting local intensity. The cross-cohort design treats pandemic exposure as mainly determined by birth timing, comparing cohorts in utero during the pandemic's peak with adjacent cohorts, for example at annual or quarter-of-birth level (Almond 2006; Fletcher 2018; Mazumder et al. 2010; Myrskylä, Mehta and Chang 2013). A main limitation of cohort comparisons is that they require an exchangeability assumption (Beach et al. 2022), which is the 1919 birth cohort (exposed in utero to the 1918 pandemic) and surrounding birth cohorts (not exposed) are similar in all relevant aspects except for pandemic exposure. This assumption can fail, and Beach et al. (2022) provide direct evidence on this point: they show that the 1919 birth cohort came from families of lower socioeconomic status (SES), violating the exchangeability assumption, and they also show that estimated adult SES deficits shrink and become statistically insignificant once background characteristics are controlled for. They further emphasize that the timing of the pandemic overlapped closely with World War I mobilization and deployment, creating plausible channels through which parental composition (lower SES fathers of children born during the pandemic) and maternal conditions (elevated stress among pregnant women due to wartime mobilization efforts) could differ across adjacent cohorts.

The second group exploits geographic variation in exposure intensity, typically proxied by influenza- and pneumonia-related mortality, to identify a dose-response relationship between pandemic severity and later-life outcomes. Almond (2006), in addition to cohort comparisons,

implements a dose-response approach using maternal mortality rates as a proxy for exposure intensity. Beach et al. (2022) replicate Almond's (2006) dose-response analysis and extend the approach using city-level intensity measures, finding that evidence from dose-response designs is sensitive and, in several cases, mixed. More broadly, review work by Beach, Clay and Saavedra (2022) states that "[v]ariation in pandemic severity has proven to be a useful tool for understanding the health consequences of 1918" (p. 77). However, a key limitation of dose-response designs is that raw mortality-based intensity measures may be confounded by place-specific factors that affect both pandemic mortality and long-run outcomes, such as pre-pandemic mortality patterns (Bootsma and Ferguson 2007), poverty and proximity to military bases (Clay, Lewis and Severnini 2019). On this point, Clay, Lewis and Severnini (2018) highlight that air pollution can amplify infectious-disease mortality during the 1918 pandemic, implying that observed influenza mortality differences across places can partly proxy environmental conditions rather than purely pandemic severity. This creates an identification challenge: a simple dose-response analysis could mistakenly attribute to "pandemic intensity" what is in part the effect of place characteristics correlated with both mortality in 1918 and life-course outcomes.

Taken together, this literature motivates combining timing-based cohort comparisons with intensity-based geographic variation, while explicitly accounting for selection. In this regard, Beach et al. (2022) proposes a unique measure of city-level excess influenza mortality, which arguably reduces confounding. In Section 3 we provide detailed discussion of this measure, which we also use as a key variable in our study.

C. This Study's Contributions

The contribution of this study is multi-fold. First, it extends the fetal origins literature on the 1918 influenza pandemic by shifting the focus from human capital, earnings, and midlife health outcomes to old-age mortality (longevity), which captures the ultimate long-run health consequence and is directly welfare-relevant. Second, it focuses on low-intensity cities, showing that low intensity does not imply a small fetal shock: if high-intensity places induce stronger private protective behaviors or stronger early-life selection (culling), then marginal exposure in low-intensity places could translate more directly into fetal damage and later-life scarring. In this sense, this paper uses a scarring-versus-selection framework to explain why detectable longevity effects can be concentrated in low-intensity settings, which have received limited attention in the

existing literature. Specifically, this study adds new evidence to Beach et al. (2022): they use the adjacent-cohort comparison between the 1918 and 1919 cohorts to help rule out unobserved cohort differences, but they do so in a pooled sample and do not separately analyze low- versus high-intensity cities. In their analysis, the estimated effects are statistically insignificant, which is consistent with a scarring versus selection interpretation emphasized by this study: pooling cities may wash out effects because selection may offset scarring in high-intensity cities, while weaker culling allows scarring to show up in low-intensity cities.

Empirically, this study contributes by implementing an identification strategy that combines within-city adjacent-cohort comparisons with plausibly exogenous cross-city variation in unexpected local severity, while directly addressing the central selection concern emphasized by Beach et al. (2022). The key comparison is between 1919 births (treated) and 1918 births (control), and the CenSoc-DMF (version 4.0, the most recent version at the time of this study) data provide exact birth year and month, enabling a clean adjacent-cohort design and the use of birth month fixed effects to control for seasonality in birth outcomes highlighted by Currie and Schwandt (2013). To strengthen identification beyond cohort timing alone, this study uses Beach et al.'s (2022) city-level excess influenza mortality measure (constructed to capture the unexpected 1918 mortality surge relative to each city's pre-1918 trend) and controls for an extensive set of parental and household background characteristics. Furthermore, the assembled linked-census dataset also allows individuals to be followed from childhood (1920) to young adulthood (1940) and midlife (1950), enabling more direct mechanism tests and shedding lights on how early-life pandemic exposure could translate into later-life disadvantage, possibly through labor-market entry job quality and persistent wage trajectories, rather than relying on indirect proxies such as observing individuals only once at one later point in adulthood about their income or education outcomes. Methodologically, this paper also demonstrates how causal ML (e.g., generalized random forest and double/debiased ML) can uncover heterogeneity that conventional subgroup analyses may miss, with parental education possibly being an important moderator of the longevity impact.

III. Empirical Strategy

A. Data

CenSoc-DMF.—The main data source of this study is the SSA's DMF linked to the full-count 1940 U.S. Census (CenSoc-DMF), which was designed for studying old-age mortality in the United States.[9] It links males observed in the 1940 Census to the SSA's DMF and covers deaths occurring at ages 65 or above (Breen, Osborne and Goldstein 2023). It also includes a unique identifier that allows researchers to merge mortality outcomes back onto individual-level variables in the publicly accessible 1940 Census data provided by IPUMS USA.[10] A key advantage of CenSoc-DMF is its scale: the most recent version (i.e., version 4.0) used by our study contains about 5.3 million deaths that occurred between 1975 and 2020, far exceeding typical survey-based datasets on aging and making heterogeneity analyses more feasible than in surveys such as the Health and Retirement Study.

CenSoc-DMF was constructed using a deterministic record-linkage algorithm described in detail by Breen and Osborne (2022). It includes males only, because changes of females' surnames after marriage make linkage infeasible, and the DMF also lacks parents' surnames and maiden-name information, which prevents linking females who changed their maiden names.[11] The mortality-adjusted match rate[12] is about 30%, and CenSoc-DMF is broadly representative: it is slightly skewed toward higher socioeconomic status (e.g., higher high-school completion than the general population), and Black males are under-represented in the linked DMF sample (Breen and Osborne 2022).

---

[9] For details, see https://censoc.berkeley.edu/data/ (accessed in September 2025).

[10] IPUMS USA (originally, the "Integrated Public Use Microdata Series") is a public data project that provides access to integrated and high-precision microdata including the U.S. decennial censuses (source: https://usa.ipums.org/usa/, accessed in September 2025).

[11] In one of the robustness checks, this study also uses the CenSoc-Numident data (Goldstein et al. 2023), which include SSA death records of both males and females linked with the full-count 1940 U.S. Census. However, the females in the linked sample are only those whose surnames did not change (for example, those who were never-married) and who can be linked using their fathers' surnames (Breen, Osborne and Goldstein 2023). A further limitation of the CenSoc-Numident data is the shorter death coverage window (1988–2005), which is substantially narrower than that of the CenSoc-DMF version 4.0 (1975–2020). This shorter window makes truncation more likely, lowering the ability to observe old-age mortality for more recent cohorts. We discuss the consequence of this truncation problem and CenSoc-Numident in Appendix B.

[12] This is calculated as (a)×(b), where (a) is the proportion of individuals from the 1940 Census successfully matched to SSA's DMF, and (b) is the inverse probability of dying in the mortality observation window conditional on living till 1940 (Breen and Osborne 2022).

One point worth emphasizing is that this study uses the most recently released CenSoc-DMF (version 4.0) dataset, which covers mortality for the years 1975–2020. Given that our study focuses on the 1918 and 1919 birth cohorts, the length of this death coverage window should substantially reduce truncation and increase observable old-age mortality (i.e., death at age 65 or above). Indeed, the birth cohorts analyzed in this study fall within CenSoc-DMF's recommended range (1900–1930), for which old-age mortality is well captured in CenSoc-DMF.

Another important feature is that CenSoc-DMF provides exact birth years and exact birth months (Breen, Osborne and Goldstein 2023). This allows our study to define adjacent cohorts cleanly (e.g., 1919 vs. 1918) and to control for seasonality in birth outcomes emphasized by Currie and Schwandt (2013) using birth month fixed effects. In contrast, census data do not directly record birth year; birth year can only be inferred from reported age in the census, which introduces measurement error.[13]

IPUMS USA Census Data.—This study also uses the 1920 full-count census, the 1920–1940 linked census, and the 1940–1950 linked census provided by IPUMS USA (Ruggles, Flood, Sobek et al. 2025; Ruggles, Nelson, Sobek et al. 2024; Ruggles, Ozder, Fitch et al. 2025).[14] The 1920 full-count census is a complete individual-level enumeration of the U.S. population in 1920. It provides detailed individual-level and household-level information when the key cohorts in this study (e.g., those born in 1918 and 1919) were observed as children living with parents, providing important background characteristics which we use as control variables.

The 1920–1940 linked census and the most recently released 1940–1950 linked census come from the IPUMS Multigenerational Longitudinal Panel project,[15] which allows the same

[13] In the census data available at the IPUMS USA (information provided at this website: https://usa.ipums.org/usa-action/variables/BIRTHYR#comparability_section, accessed in September 2025): "self-reported birth year is not available but has been calculated by subtracting AGE from the survey year." A key implication is that this calculated birth year is off by one year for everyone in Census Year X who has a birthday that falls after the census reference date, which is April 1 of Year X. For example, Person A born in January 1990 will be recorded as age 10 on the 2000 Census (as of April 1, 2000), so the calculated birth year ($2000 - 10 = 1990$) is correct. In contrast, Person B born in December 1990 will be recorded as age 9 on the 2000 Census (because December 2000 is after April 1, 2000), so the calculated year of birth ($2000 - 9 = 1991$) is off by one year. Avoiding this issue requires information on birth month. However, birth month information in the census data is only available in the 1950, 1910, 1900, 1880 and 1870 samples (for more details, see https://usa.ipums.org/usa-action/variables/BIRTHMO#availability_section, accessed in September 2025).

[14] Data are available at https://usa.ipums.org/usa-action/variables/group (accessed in September 2025). This study focuses on those born within the United States. In the estimation sample, there are no data from Alaska and Hawaii.

[15] More detailed information is provided at https://usa.ipums.org/usa/mlp/mlp_extracts.shtml (accessed in September 2025).

individual to be followed across censuses. Using these linked data, our study is able to observe childhood conditions in 1920, adult outcomes in 1940, and later-life socioeconomic status in 1950 for the same person. This structure enables direct tests of long-run pathways: for example, whether in utero exposure affects early adult human capital and labor-market entry conditions and whether those impacts persist into midlife, analyses that are typically infeasible without person-level links across multiple censuses.

Beach et al. (2022) Data.—Beach et al. (2022) construct a city-level pandemic intensity measure using influenza death data for U.S. cities provided in the Census Bureau's *Mortality Statistics* reports. The key contribution of this measure is that it moves beyond crude state-level proxies and instead captures local variation across nearly 300 cities, a geographic level at which the 1918 influenza shock could have exogenous variation in severity. What makes this measure essential to our study is that it captures *excess* influenza mortality relative to a city-specific counterfactual, rather than raw influenza mortality.

To construct this measure, Beach et al. (2022) transcribe city-year mortality data for 1900–1930, and then for each city estimate a pre-pandemic trend using the 1900–1917 period, and then predict what influenza deaths in 1918 would have been in the absence of the pandemic. Pandemic intensity is then defined as the unexpected increase in 1918 influenza deaths, which is actual influenza deaths in 1918 minus predicted influenza deaths in 1918. Beach et al. (2022) further normalize this unexpected increase by predicted 1918 influenza deaths, generating a relative shock measure rather than one driven mechanically by city size. This approach is important because it addresses a key concern emphasized by Beach et al. (2022): places with higher influenza mortality in 1918 also differed in environments, such as air pollution, which could affect long-run outcomes. By comparing each city's 1918 influenza deaths to what that same city would have been expected to experience based on its earlier trend, the measure constructed by Beach et al. (2022) captures excess deaths caused by the pandemic, rather than a city's usual illness level. This helps remove differences across cities that existed even before 1918, such as persistently poor health conditions or long-run mortality patterns.

This study uses the measure constructed by Beach et al. (2022) to exploit plausibly exogenous city-level variation in pandemic intensity, that is, variation that is unlikely to be correlated with cohort composition or parental characteristics. Specifically, our study uses the median value of this measure to classify cities as low- or high-intensity. This classification is based on a shock measure

explicitly designed to capture unexpected local severity in 1918, rather than broader city attributes that could confound exposure-mortality relationships. Figure 1 maps the cities included in the estimation sample of our study. Figure 2 maps these cities by pandemic intensity. Overall, the coverage is broad: cities are from major U.S. regions as of the 1920s (Northeast, Midwest, South, and West), with denser coverage in the Northeast and Midwest, sparser coverage in parts of the South and West, and no coverage of Alaska or Hawaii.

B. Identification Strategy

This study's identification strategy combines within-city adjacent-cohort comparisons with cross-city variation in pandemic intensity, while addressing the central concern emphasized by Beach et al. (2022). The key comparison is between 1919 births (in utero during the 1918 pandemic), which form the treatment group, and 1918 births, which form the control group. Importantly, the exact birth year information in the CenSoc-DMF data enables a clean adjacent-cohort design, improving on prior cross-cohort studies that must infer birth year from census-reported age. Beach et al. (2022) emphasize that a simple comparison of the 1919 cohort and adjacent cohorts is confounded because cohorts differ in family background due to selection into births. To address this concern, we follow the identification strategy advocated by Beach et al. (2022), which is using a localized excess influenza mortality measure that captures the unexpected 1918 mortality surge relative to each city's pre-1918 trend and is plausibly exogenous.

Specifically, we use the merged linked-census dataset and include the same set of background characteristics used by Beach et al. (2022) and deemed as important confounding factors. We provide evidence that these background characteristics are balanced between the 1919 and 1918 cohorts conditional on city-level pandemic intensity, and estimated effects are robust to including background covariates along with city and birth month fixed effects. Together, these results suggest that the estimated longevity differences may not be confounded by differences in background or other baseline characteristics, rather than fetal scarring.

It is also worth emphasizing that, consistent with Beach et al. (2022), this paper relies on a city-level excess influenza mortality measure constructed to capture the unexpected 1918 influenza mortality surge relative to each city's pre-1918 trend. Figure 2 shows that cities are dispersed without clear geographic concentration: low- and high-intensity cities often coexist within the same broader area. This spatial pattern is consistent with plausibly exogenous variation in

unexpected 1918 severity across cities. The linked census data further allow our study to expand the mechanism checks conducted by Beach et al. (2022). Following the same individuals across 1920–1950 and into old-age mortality outcomes enables more direct examination of pathways, such as linking early-life exposure to labor-market entry job quality and persistent wage trajectories, rather than relying on indirect proxies.

Our study also complements the main design with additional credibility checks, including falsification tests using other adjacent cohorts, Lee bounds (Lee 2009) for sample selection, Oster's sensitivity analysis (Oster 2019) for selection on unobservables, and robustness checks on alternative intensity definitions. We also use causal ML tools (e.g., generalized random forest and double/debiased ML) to characterize heterogeneity in a methodologically disciplined way in settings where conventional subgroup splits are under-powered.

A related concern about our research design is whether record linkage could generate spurious exposure-longevity correlations. This is unlikely here because both the DMF-census and cross-census links are determined mainly by name information (and related identifiers used for record matching), rather than by characteristics that directly reflect an individual's health. As a result, linkage rules are unlikely to mechanically generate or confound the estimated relationship between in utero pandemic exposure and later-life longevity. To address concerns about attrition, we apply the bounding method proposed by Lee (2009), following Lin and Liu (2014).

Replication of Key Results of Beach et al. (2022).—Our identification strategy largely follows Beach et al. (2022). While our study, whenever possible, uses the same set of variables constructed in the same way as Beach et al. (2022), the estimation sample differs from theirs. In particular and as noted previously, Black males are under-represented in the linked DMF sample (Breen and Osborne 2022). In addition, the dataset used by our study, which is CenSoc-DMF merged with linked census data, is a subset of the data used by Beach et al. (2022). This raises concerns about positive selection, since only those who survive to age 65 are observed in our sample. Despite these differences in sample construction, data sources, and the potential for positive selection, focusing on the same set of variables, our study reproduces the exact same key pattern reported by Beach et al. (2022) in their figure 2. Our results are reported in Figure 3. Specifically, adjacent birth cohorts are not mechanically comparable because family background shifts around the 1919 cohort: paternal socioeconomic proxies for the 1919 birth cohort and own schooling deviate from smooth cohort trends (likely due to World War I enlistment, as discussed by Beach et al. 2022),

indicating non-random selection into births. This implies that a simple 1919 versus 1918 comparison can be confounded if selection is not addressed. By showing that lower socioeconomic status (proxied by years of education) can be explained by differences in family background (proxied by paternal characteristics), our study cross-validates the central conclusion of Beach et al. (2022): controlling for parental and household characteristics is essential, and without doing so, the estimated relationship between pandemic exposure and later-life outcome reflects negative selection of the 1919 cohort's parents, as indicated by Figure 3.[16]

Balance of the Background Characteristics.—Figures 4 and 5 of our study further cross-validate the findings of Beach et al. (2022) reported in their table 2 and table 6 respectively, using the same set of background characteristics examined by Beach et al. (2022). Specifically, Figure 4 of our study confirms the pattern reported by Beach et al. (2022):[17] relative to fathers of adjacent cohorts, fathers of the 1919 birth cohort have significantly lower socioeconomic status. They are more likely to be illiterate, have a lower Duncan Socioeconomic Index (SEI) score, work in occupations with lower occupation income scores, and be born outside the United States. The 1919 cohort is also slightly more likely to live in a family-owned home, slightly more likely to be born in the South, and to have older fathers and more older siblings. Figure 5 also confirms the results reported by Beach et al. (2022), showing that background characteristics examined in Figure 4 are balanced, except for father's age at birth being different by about 0.2 years, once conditioning on the standardized city-level excess influenza mortality measure, birth year fixed effects, and residential city in 1920 fixed effects:[18] differences in those background characteristics are statistically insignificant, and their point estimates are close to zero. The balance results are also consistent with our study's finding that estimates of the in utero exposure effects on old-age mortality are similar with or without controlling for these background characteristics in the regression model once we condition on city-level excess influenza mortality, which we discuss in the next section and in the results section.

[16] We set 1914 as the starting birth cohort. Individuals born in 1914 were younger than six during the pandemic and are therefore unlikely to have experienced pandemic-related disruptions to schooling.
[17] The use of robust standard errors follows Beach et al. (2022).
[18] Here, the standardization follows Beach et al. (2022), which is defined as city-level excess influenza deaths (provided by Beach et al. 2022) divided by the sample mean. This standardization allows the coefficient of "Born in 1919 (1/0) × Standardized Excess Flu" to be interpreted as the average difference in the background characteristics between the 1919 birth cohort and adjacent birth cohorts.

C. Econometric Specification

The empirical analysis of our study starts with the following regression model, which follows Beach et al. (2022):

$$y_{ibc} = \alpha_c + \beta_b + \delta 1\{YOB = 1919\} \times Flu_c + control_i + \epsilon_{ibc}, \quad (1)$$

where y_{ibc} denotes outcome y for individual i born in year b and living in city c in 1920. The parameters α_c and β_b are fixed effects of residential city in 1920 and birth year, respectively. Flu_c denotes the standardized city-level excess influenza mortality measure. This standardized measure is defined as the city-level excess influenza deaths constructed by Beach et al. (2022) divided by the sample mean.[19] The vector $control_i$ includes control variables measured in the 1920 census and used for individual i, which we refer to as background characteristics in this paper. This is the same set of background characteristics used by Beach et al. (2022): dummy variables for own race, own state of birth, father's state of birth (if within the U.S.) or country of birth (if outside the U.S.) of birth, mother's state of birth (if within the U.S.) or country of birth (if outside the U.S.), father's occupation code (based on the Census Bureau occupational classification system), a second-generation immigrant father, whether the father moved outside of birthplace before child was born, an illiterate father, an illiterate mother, whether the mother was in the labor force, whether the family owned the home; father's age when child was born; mother's age when child was born; the difference between father's and mother's ages; and birth order.

To examine impacts by pandemic intensity, we use the median of Flu_c to define low-intensity cities (less than or equal to the median) and high-intensity cities (greater than the median).[20] In this case, Flu_c is replaced by a binary indicator of high intensity (or low intensity), and in estimations restricted to the 1918 and 1919 birth cohorts, β_b captures the effect of in utero exposure to the 1918 influenza pandemic on y (e.g., old-age mortality) in low-intensity (or high-intensity) cities.[21] In the most inclusive specification, equation (1) also includes birth month fixed

[19] This is the same standardization used by Beach et al. (2022), and this standardization allows the coefficient of that interaction term to be interpreted as the average effect of pandemic exposure.

[20] We use the median because it is less sensitive to extreme values (if any) in the standardized city-level excess influenza mortality measure. In addition, a median split generates two approximately equal-sized groups, which facilitates statistical comparisons and helps keep statistical power more balanced across groups. This approach is also consistent with the review article done by Beach, Clay and Saavedra (2022) about the 1918 influenza pandemic, who use below-median and above-median for their by-pandemic-intensity analysis (reported in their figure 6).

[21] The reason why we do not use subsample estimations by pandemic intensity is to avoid using small samples that may lack sufficient statistical power.

effects. All estimations are conducted using ordinary least squares (OLS), with standard errors clustered at the residential city in 1920 level.

Summary statistics based on our study's estimation sample are reported in Table 1. We observe that the average age at death is similar between low-intensity and high-intensity cities, but it is important to point out that this unconditional comparison masks adverse effects of pandemic exposure, which our study identifies through within-city adjacent-cohort comparisons (discussed in the next section).

IV. Results

A. Main Results

Estimations Using Adjacent Birth Cohorts.—Figure 6 presents estimates of the impact of pandemic exposure on old-age mortality based on equation (1), which we extend by adding interaction terms between indicators of birth cohorts other than 1919 and the standardized city-level excess influenza mortality measure with birth year 1917 being used as the omitted period in the regression model. Relative to this omitted group, the estimates suggest a downward pattern in age at death for the 1918 and 1919 cohorts. This pattern, although not statistically significant, is consistent with a potential long-run impact of pandemic exposure and with effects that increase with the duration of exposure: the 1919 cohort could have been exposed throughout gestation, whereas in the 1918 cohort only those born in the last quarter of 1918 may have been exposed during the final trimester. Also notably, the point estimates remain similar whether or not background characteristics are controlled for, which is consistent with the possibility that exposure becomes exogenous once conditioning on the city-level excess influenza mortality measure.

Note that results reported in Figure 6 indicate the old-age mortality impact of pandemic exposure in cities and evaluated at the mean level of pandemic intensity. Next, we examine the impact separately for low-intensity and high-intensity cities; these results are reported in Figure 7. In all OLS estimations conducted in Figure 7, we control for background characteristics, residential city in 1920 fixed effects, and birth month fixed effects. In low-intensity cities, the focal comparison—1919 births (in utero during the pandemic) vs. 1918 births—shows a clearly negative effect: the point estimate is about -0.25 years (roughly three months) and is statistically significant. This result suggests that scarring is detectable even where pandemic intensity was modest. In

contrast, the placebo adjacent-cohort estimated differences (1918 vs. 1917, 1917 vs. 1916, 1916 vs. 1915, and 1915 vs. 1914) cluster near zero with confidence intervals that include zero. These comparisons serve as falsification tests and indicate that the main finding is unlikely to reflect a cohort trend in longevity.

In high-intensity cities, the corresponding 1919 vs. 1918 estimate is smaller in magnitude (around -0.13 years) and statistically indistinguishable from zero. Differences in age at death estimated using other adjacent-cohort comparisons are also statistically insignificant. This pattern is consistent with the hypothesis that selection (culling) may offset scarring when the shock is severe. It is also important to note that the estimated exposure effect in high-intensity cities is imprecise, meaning that the damaging effect of pandemic exposure in high-intensity cities could be mixed with the selection (culling) effect. Furthermore, the overlap in confidence intervals between the effect in low-intensity cities and the effect in high-intensity cities implies that the two effects could be equal. In this regard, our study does not claim that there is no effect of pandemic exposure in high-intensity cities; rather, it is the possible presence of selection effect that can make the estimated effect highly imprecise. Overall, Figure 7 indicates that a specific longevity deficit of the 1919 cohort relative to the 1918 cohort emerges where fetal scarring may dominate (in low-intensity cities), and such deficits seem to be absent in other adjacent-cohort comparisons or in high-intensity cities.

Results in Figure 7 are based on the regression model that uses the most inclusive specification. Table 2 and Table 3[22] report more detailed results of alternative specifications for low-intensity and high-intensity cities, respectively. Overall, estimates of the longevity effect of pandemic exposure are robust to alternative specifications; in particular, the point estimates are similar with and without controls for background characteristics, which is consistent with the balance-check results discussed earlier.

Estimations by Birth Order Using the 1919 and 1918 Cohorts.—Figure 8 repeats the estimations in Figure 7 using subsamples by birth order and focusing on the 1919 and 1918 cohorts. In low-intensity cities, exposed first-born and high-parity (5+) males experience longevity losses of about 0.572 years (roughly seven months) and 0.331 years (roughly four months), respectively,

[22] The 1918 birth cohort may be partially exposed to the 1918 influenza pandemic, particularly those born in the last quarter of 1918. In that case, the estimated difference in old-age mortality between the 1918 and 1917 birth cohorts could reflect this partial exposure operating through a scarring effect.

whereas exposed 2nd–4th born males do not show comparable losses. In high-intensity cities, only first-born males exhibit a statistically significant longevity loss of about 0.497 years (roughly six months), while estimates for later-born males are close to zero and statistically insignificant.

Overall, Figure 8 suggests a U-shaped baseline vulnerability by parity combined with parity-specific selection. One explanation for this U-shaped pattern could be health disparities over the life course: first-borns have worse health at birth, but during adolescence the birth-order effects switch direction due to health issues related to younger siblings engaging in more risky behaviors (Björkegren and Svaleryd 2023). Another explanation could be maternal age: first births are more likely to occur to younger mothers, whereas later-parity births are more likely to occur to older mothers. Myrskylä and Fenelon (2012) show that offspring of mothers younger than 25 or older than 35 have worse adult outcomes (including mortality) than offspring of mothers aged 25–34. Based on the scarring versus selection framework, Figure 8 suggests that when the pandemic is less severe, scarring among already-fragile first-born and high-parity pregnancies could be detectable; when the pandemic is severe, later-born males may be so strongly culled that surviving high-parity males are positively selected and therefore no longer exhibit scarring in old age, while first-born males continue to show substantial damage.

Estimations by Birth Quarter Using the 1919 and 1918 Cohorts.—This analysis examines whether the longevity loss of the 1919 cohort relative to the 1918 cohort aligns with the timing of fetal exposure. We estimate effects separately by birth quarter, and the results are reported in Figure 9. In low-intensity cities, the quarter pattern is not monotone: the point estimate is negative for Q1, strongly negative for Q3 and Q4, and statistically insignificant for Q2.

This timing evidence is consistent with the interpretation that the low-intensity-city impacts are unlikely to be driven mainly by high-virulence maternal infection during the peak fall wave of the pandemic. Instead, the adverse effects also appearing for Q3 and Q4 1919 births suggest that, in low-intensity setting, limited exposure to virulence could lead to limited fetal culling (selection), allowing scarring among survivors to remain visible in old-age mortality, potentially through broader pathways (e.g., stress and disrupted household resources) that need not be concentrated only among Q1 1919 births. One possible explanation for the Q1 effect is that, because the fall wave of the 1918 influenza pandemic was the most severe phase of the outbreak in the United States, those born in Q1 of 1919 could have been exposed to the fall wave during the second trimester; Helgertz and Bengtsson (2019) find that for males (which is the focal population of our

study), second-trimester exposure could have impact on later-life hospitalization risk. One possible explanation for the lack of an effect for Q2 is that those born in Q2 of 1919 could have been exposed to the fall wave during the first trimester; Ogasawara (2022) finds that first-trimester exposure decreases the proportion of males at birth, suggesting fetal culling (selection); Helgertz and Bengtsson (2019) find that for socioeconomic outcomes (e.g., income and occupation), first-trimester exposure can be positively associated, likely reflecting selection effects rather than a direct benefit of exposure. In high-intensity cities, Figure 9 shows that the estimated longevity loss is close to zero and statistically insignificant for each birth quarter. This pattern is also consistent with the possible presence of selection (culling) effects.

B. Robustness Checks

This section briefly summarizes our robustness checks. Appendix A provides more detailed discussion. The main results on the effects of pandemic exposure on old-age mortality (Figure 7) are robust to the several alternative specifications: defining low- and high-intensity cities using the sample mean rather than the sample median of the standardized city-level excess influenza mortality measure (explained in the empirical strategy section); basing inference on standard errors clustered at the county (instead of city) level; and using the median of the standardized city-level excess influenza mortality measure computed from the original data provided by Beach et al. (2022), rather than the median computed from our study's estimation sample. Appendix Table A1 and Appendix Table A2 report the results of these robustness checks for low- and high-intensity cities (focusing on the 1919 and 1918 birth cohorts), respectively.

Following Lin and Liu (2014) whose study is comparable to ours, and applying Lee bounds (Lee 2009), we find that excess attribution in the treatment group (i.e., the 1919 birth cohort) relative to the control group (i.e., the 1918 birth cohort), potentially due to poorer health, could cause the original estimates to understate the true impact of pandemic exposure. Detailed results are reported in Appendix Table A3.

Using the sensitivity analysis proposed by Oster (2019), we find that the key estimate of the scarring effect—the impact of pandemic exposure on old-age mortality in low-intensity cities—remains robust when allowing for selection on unobservables. Detailed results are reported in Appendix Figure A1.

Furthermore, we conduct a difference-in-differences (DID) analysis to remove the impact of any potential "one-year younger" survival disadvantage (e.g., needing to live longer to reach age 65). Specifically, we compare the 1919 versus 1918 longevity gap with the analogous gap for an earlier adjacent-cohort pair (i.e., 1915 versus 1914). Detailed results are reported in Appendix Tables A6 and A7. Overall, the DID estimates are similar to the main results reported in Tables 2 and 3, and they also remain stable with background controls and various fixed effects, supporting scarring effects in low-intensity settings. In high-intensity cities, the DID estimates are statistically insignificant throughout, consistent with selection offsetting scarring. Placebo DID tests using earlier cohorts show no effects in either low-intensity or high-intensity settings.

C. Intensity Split Validation and Potential Mechanisms

Intensity Split Validation.—In this section we first test whether our low- versus high-intensity city classification is driven by pandemic intensity itself, rather than by other correlated city characteristics. In other words, these checks help rule out the possibility that "low" and "high" simply proxy for unobserved city-level heterogeneity. Results are reported in Figure 10.

In Case 1, low-intensity cities in the estimation sample are those located in Michigan, Minesota and Wisconsin—the bottom three states in the 1918 influenza mortality rate ranking reported in table 1 of Garrett (2008); high-intensity cities are those in Pennsylvania, Maryland and New Jersey—the top three states in the same ranking. In this case when "low" and "high" are defined using an external severity ranking, the pattern still lines up with intensity: the estimate of the 1919 versus 1918 age-at-death gap for the low-intensity group (about -0.282 years) is similar to the main low-intensity estimate reported in Figure 7 (about -0.250 years), while the effect for the high-intensity group is statistically insignificant. Because this split relies on an external measure of severity, it supports the claim that our classification captures pandemic intensity rather than city-level unobserved heterogeneity.[23]

[23] Appendix Table A4 reports the results for low-intensity cities in the estimation sample that are in the bottom 10 states based on the 1918 influenza mortality rate ranking reported in table 1 of Garrett (2008); results are similar to those reported in Table 2 (the main results). Appendix Table A5 reports the results for high-intensity cities in the estimation sample that are in the top 10 states based on the 1918 influenza mortality rate ranking reported in table 1 of table 1 of Garrett (2008); results are similar to those reported in Table 3 (the main results); here, the lack of significance in the 1918 versus 1917 case could be explained by mortality measured in 1918, which did not affect those born in 1918 but in utero during 1917. In addition, results in Appendix Table A4 and Appendix Table A5 are similar to the results shown in Figure 10 Case 1. We do not conduct the estimations of top 10 cities versus bottom 10 cities, to avoid small samples. Although there is a ratio of 1918 and 1915 mortality rates reported in table 1 of Garrett

In Case 2, we restrict the sample to states that contain both low-intensity and high-intensity cities. This restriction reduces the risk that the low versus high comparison simply reflects state-level differences in, for example, institutions, health systems or economic conditions. Based on this subsample, the low-intensity estimate remains essentially the same as the main result (-0.25 years, reported in Panel A column 4 of Table 2), suggesting that the impact in low pandemic intensity areas are not confined to specific states. In contrast, the estimated impact in high pandemic intensity areas remains statistically insignificant and similar to the main result (reported in Panel A column 4 of Table 3).

In Case 3, we further restrict the Case 2 sample to states in the Northeast (PA, NY, NJ, CT, RI, MA, NH, VT and ME), where socioeconomic conditions may be more comparable across states. The results remain very similar to those in Case 2, again supporting the interpretation that the "low-intensity" versus "high-intensity" split is intensity-driven. The similarities between the results of Case 2 and Case 3 and the main results (reported in Tables 2 and 3) are also consistent with prior literature finding influenza mortality rates to be uncorrelated with state economic conditions, climate, or geography (Garrett 2008).

In Case 4, we restrict the sample to individuals living in rural areas in 1920 within the six states used in Case 1.[24] In this rural subsample, we find that the estimates remain negative for the bottom three states and the top three states from Case 1, but they are statistically insignificant. In rural areas, there may have been less use of quarantines, implying less stress and therefore smaller impacts. Overall, Case 4 suggests that the main results (reported in Table 2's Panel A and Table 3's Panel A) are driven by cities rather than by rural areas within the states that contain those cities.

To further assess whether the main results are driven by cities, in Case 5 we conduct estimations using Philadelphia, PA (the residential city in 1920) as the high-intensity city and New York City, New York (the residential city in 1920) as the low-intensity city.[25] Overall, this two-city case study reproduces the main results of this study previously discussed. The estimated impact of pandemic exposure remains statistically significant and larger in magnitude in the low-intensity city,

(2008), we do not use it in order to reduce the impact of potential measurement error. This is because people may react to the absolute level of mortality, rather than a relative measure, and they may not even know what the 1915 mortality rate was.

[24] In this case, residential city in 1920 fixed effects are replaced by residential county in 1920 fixed effects; standard errors are clustered at the residential county in 1920 level.

[25] In this case, residential city in 1920 fixed effects are no longer used; robust (to heteroskedasticity) standard errors are used to construct confidence intervals.

consistent with scarring. In contrast, the impact is much more noisily estimated and statistically insignificant in the high-intensity city, consistent with selection (culling) offsetting scarring.

We choose these two large cities to ensure sufficient statistical power. More importantly, several plausible confounders that might otherwise explain the different pattern in the impact of pandemic exposure on old-age mortality between low-intensity and high-intensity settings were broadly similar for Philadelphia and New York City during the 1918 pandemic. First, maternal stress from husbands' World War I military deployment is unlikely to account for the divergence: based on the number of soldiers (excluding officers) supplied by each state, New York and Pennsylvania ranked first and second, respectively (Ayres 1919, 23). Second, both cities were major early-20th-century industrial centers and plausibly exposed to significant air pollution from industrial coal-fired plants—a confounder of the impact of pandemic on mortality pointed out by Clay, Lewis and Severnini (2018). In this context, the Philadelphia versus New York City contrast is not a "clean air versus dirty air" comparison. Third, wartime increases in food prices were a common shock, and both cities faced high prices, so differential food inflation (triggering reduced food intake during pregnancy) is unlikely to explain the pattern.[26]

A key difference between the two cities during the pandemic was the public-health policy response. Philadelphia responded late, implementing nonpharmaceutical interventions after mortality had already accelerated, and experienced a high mortality rate. New York City acted early, enforcing sustained measures such as isolation and quarantine before the pandemic escalated, and achieved "the lowest death rate on the Eastern seaboard" (Markel et al. 2007, 651). Given the similarities between the two cities and the large difference in their pandemic mortality rates, it is reasonable to interpret the pattern in Figure 10's Case 5 as reflecting intensity rather than city-specific unobserved heterogeneities. Overall, Case 5 confirms the key findings of our study: scarring effects may be detectable at low pandemic intensity, whereas selection (culling) effects may be more relevant at high pandemic intensity.

Potential Mechanisms.—In this section we explore potential pathways underlying the main results reported in Table 2 (Panel A) and Table 3 (Panel A). Specifically, we examine pathways beyond maternal stress during pregnancy, which Aizer, Stroud and Buka (2016) show can

[26] For details, see https://fraser.stlouisfed.org/title/retail-prices-161/retail-prices-1913-december-1919-5364 (accessed in January 2026).

adversely affect a wide range of offspring outcomes. Using the linked 1920–1940 and 1940–1950 censuses, we examine whether the effects of in utero pandemic exposure show up decades later in schooling, employment, earnings and socioeconomic status. Results are reported in Table 4. Panel A indicates that effects on years of education (measured in 1940) are limited (statistically insignificant) for low-intensity cities, and small (close to zero but statistically significant) for high-intensity cities. This pattern is consistent with Beach et al. (2022), who find mixed evidence on educational attainment in a city-level dose-response analysis. Panel B shows negative and statistically significant effects on employment in 1940 of similar magnitude in low- and high-intensity cities. Panel C shows that adverse effects on socioeconomic status (indicated by lower values of the Duncan Socioeconomic Index) are larger for low-intensity cities. Panels D and E[27] show that for those exposed to the 1918 pandemic in utero in low-intensity cities, their 1940 wages were lower and remain lower in 1950. In contrast, for those exposed to the 1918 pandemic in utero in high-intensity cities, while their wages in 1940 were lower, their wages in 1950 turned out to be higher.

Overall, results reported in Table 4 point to a labor-market channel, through which fetal scarring in low-intensity cities may translate into reduced longevity. The clearest support seems to come from job quality and earnings: in 1940, the 1919 cohort appears to sort into lower-status occupations (lower by about 1.122 in Duncan Socioeconomic index in low-intensity cities) and earn less wage and salary income (by about $1,500 to $1,640 in low-intensity cities, all in 1999 dollars), indicating lower-paid jobs upon labor-market entry. Importantly, in low-intensity cities the earnings penalty persists into 1950 (of about $565 to $636, all in 1999 dollars), indicating persistently lower-wage trajectories, which could plausibly translate into shorter longevity. In high-intensity cities, the estimated impact on wage and salary income is negative in 1940 but positive in 1950 (of roughly $670 to $850), suggesting that selection (culling) among survivors could offset, or even reverse, later-life labor-market disadvantage among survivors, even if they experience early disruptions such as reduced schooling, lower employment likelihood, and lower

[27] Note that the sample sizes in Panels D and E are substantially smaller. In the 1950 Census, wage and salary income information was collected only for so-called sample-line persons (for details, please see https://usa.ipums.org/usa-action/variables/SLREC#description_section, accessed in September 2025), which explains the much smaller sample size in Panel E. To have longitudinal data, in the 1940 data of the 1940–1950 linked census we focus on the same sample-line persons in the 1950 Census. This is why the sample size in Panel D is identical to that in Panel E. Importantly, sample-line status is not based on individual characteristics. It is determined solely by the line on the census form, that is, the row on the enumeration sheet on which the person was recorded.

occupation status at labor-market entry (around age 20), as reflected in the lower Duncan Socioeconomic Index measured in 1940 for the 1919 birth cohort.

Taken together, Table 4 suggests a pathway in which fetal scarring could slightly reduce schooling and push those born in low-intensity cities into lower-quality early jobs, depressing earnings persistently, and ultimately lowering longevity, potentially through the socioeconomic status and working-conditions gradient.

D. Causal ML Evidence on Heterogeneity

Figure 11 report results using causal ML methods to address a key limitation of conventional heterogeneity analysis that relies on subgroups: small sample sizes and limited statistical power when splitting the data into subgroups.[28] Specifically, we estimate individualized average treatment effects (IATE) on age at death for the 1919 birth cohort (the treated group) compared with the 1918 birth cohort (the control group) using generalized random forest (Athey, Tibshirani and Wager 2019) and double/debiased ML with 10-fold cross-fitting (Chernozhukov et al. 2018; Kennedy 2023; Knaus 2022; Semenova and Chernozhukov 2021). Control variables in the estimation include birth month dummy variables, residential city in 1920 dummy variables, and the same set of background characteristics used in previous OLS estimations. Estimation is conducted separately for low-intensity and high-intensity cities. By estimating IATE through modeling the treatment effect as a flexible function of covariates, these ML tools provide a methodologically disciplined way to characterize heterogeneity that is difficult to detect using traditional split-the-sample approaches.[29]

The main idea of such characterization method is checking whether individuals' characteristics are different between individuals that have high predicted IATEs and those that have low predicted IATEs. Such differences suggest that these characteristics could be correlated with the treatment effect, that is, treatment effect heterogeneity along those dimensions. Results reported in Figure 11 are based on sorting individuals by their estimated IATEs and partitioning them into quarters.

[28] Our estimation sample consists predominately of individuals who are White, which makes the statistical inference of subsample estimation by race significantly under-powered. Appendix Figure A2 reports such subsample estimation results, showing that estimates of the impact of pandemic exposure on old-age mortality for non-whites are very noisy with very wide confidence intervals, meaning that many null hypotheses using hypothesized values inside those confidence intervals cannot be rejected.

[29] Appendix C provides additional details on the ML methods used to estimate IATE and to construct sorted group average treatment effects based on the estimated IATE for our heterogeneous treatment effect analysis.

Means of individuals' characteristics, specifically race and education, between the top quarter and the bottom quarter are compared separately for low-intensity and high-intensity cities.[30] The plotted differences indicate a correlation between parental education and the old-age mortality impact.

In low-intensity cities, where scarring may be more likely to dominate, individuals predicted to experience greater adverse longevity impacts appear to be more likely to have illiterate parents. This pattern is consistent with the idea that lower parental human capital and fewer household resources can amplify the persistence of early-life health shocks into old-age survival. In high-intensity cities, where selection (culling) may be more relevant, the pattern appears different: individuals with greater estimated longevity impacts could come from better educated families. Given that the main effect for high-intensity cities is statistically insignificant (reported in Figure 7), we interpret the high-intensity cities' results reported in Figure 11 with caution: if anything, these results suggest that there is no clear evidence indicating that more-educated parents systematically avoided high-intensity locations during the pandemic. Overall, Figure 11 suggests that parental education could be a potentially important dimension along which the long-run mortality consequences of early-life pandemic exposure may differ. In contrast, evidence of heterogeneity by race appears limited.

V. Conclusion

In this paper we examine the long-run mortality consequences of in utero exposure to the 1918 influenza pandemic using linked individual-level census and mortality records and a city-level pandemic-intensity measure that captures the unexpected spike in influenza mortality (Beach et al. 2022). Comparing the 1919 birth cohort (in utero during the pandemic) with the 1918 cohort, we find a clear longevity deficit in low-intensity cities: exposed males live about three months less, an effect that is meaningful in magnitude and comparable to the longevity gradient associated with one year of schooling documented in prior work (Halpern-Manners et al. 2020). In high-intensity cities, the estimated effect is smaller and statistically insignificant, which should not be interpreted as no effect but is consistent with selection offsetting scarring.

[30] Here, we multiply IATE estimates by -1, given that more negative values indicate greater effects on longevity.

This cross-intensity contrast also emphasizes this study's main point: low intensity does not mean small fetal risk. In low-intensity settings, weaker fetal or infant culling can leave more scarred but surviving individuals, making old-age mortality effects detectable. In high-intensity settings, stronger early-life selection may leave a healthier group of survivors and attenuate net effects observed later in life. More broadly, the findings of our study align with the fetal-origins hypothesis that adverse prenatal conditions can have persistent impacts on later-life health and survival (Almond and Currie 2011; Barker 1995).

Results of mechanism and heterogeneity analyses indicate socioeconomic pathways: exposed males born in low-intensity cities sort into lower-quality early jobs and experience persistently lower earnings into midlife. Estimates of the pandemic's longevity effects by birth order and birth quarter, together with causal ML estimates, suggest that impacts are larger for more vulnerable births and differ by family resources, especially parental education, which is consistent with evidence that prenatal conditions, such as maternal stress during pregnancy, can have lasting effects (Aizer, Stroud and Buka 2016).

One important policy implication is that even modest-severity settings can generate long-run fetal harm, potentially through stress and other socioeconomic channels. This emphasizes the importance of policies that protect and support pregnant women during outbreaks, not only in the hardest-hit places. Finally, because our analysis relies on linked census-mortality records that cover only males observed into older ages, extending the evidence to additional populations and exploring additional pathways are important direction of future research as linked data improve.

VI. Declaration of Generative AI and AI-Assisted Technologies in the Manuscript Preparation Process

During the preparation of this work the author used ChatGPT in order to improve the readability and language of the manuscript. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the paper.

REFERENCES

Aizer, A., L. Stroud and S. Buka (2016). "Maternal Stress and Child Outcomes: Evidence from Siblings." *Journal of Human Resources* 51 (3): 523–555.

Almond, D. (2006). "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population." *Journal of Political Economy* 114 (4): 672–712.

Almond, D. and J. Currie (2011). "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives* 25 (3): 153–172.

Athey, S., J. Tibshirani and S. Wager (2019). "Generalized Random Forests." *The Annals of Statistics* 47 (2): 1148–1178.

Ayres, Leonard P. (1919). "The War with Germany: A Statistical Summary." *United States Army: Publications* 368. Retrieved from https://digitalcommons.unl.edu/usarmyresearch/368 in November 2025.

Barker, D. J. P. (1995). "Fetal Origins of Coronary Heart Disease." *BMJ: British Medical Journal* 311: 171–174.

Beach, B., R. Brown, J. Ferrie, M. Saavedra and D. Thomas (2022). "Reevaluating the Long-Term Impact of In Utero Exposure to the 1918 Influenza Pandemic." *Journal of Political Economy* 130 (7): 1963–1990.

Beach, B., K. Clay and M. Saavedra (2022). "The 1918 Influenza Pandemic and Its Lessons for COVID-19." *Journal of Economic Literature* 60 (1): 41–84.

Björkegren, E. and H. Svaleryd (2023). "Birth Order and Health Disparities throughout the Life Course." *Social Science & Medicine* 318: 115605.

Bootsma, M. C. J. and N. M. Ferguson (2007). "The Effect of Public Health Measures on the 1918 Influenza Pandemic in U.S. Cities." *Proceedings of the National Academy of Sciences* 104 (18): 7588–7593.

Bozzoli, C., A. Deaton and C. Quintana-Domeque (2009). "Adult Height and Childhood Disease." *Demography* 46 (4): 647–669.

Breen, C. F. and M. Osborne (2022). "An Assessment of CenSoc Match Quality." Retrieved from https://censoc.berkeley.edu/wp-content/uploads/2022/06/censoc_match_quality_report_2_1.pdf in September 2025.

Breen, C. F., M. Osborne and J. R. Goldstein (2023). "CenSoc: Public Linked Administrative Mortality Records for Individual-Level Research." *Scientific Data* 10: 802.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey and J. Robins (2018). "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal* 21 (1): C1–C68.

Clay, K., J. Lewis and E. Severnini (2018). "Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic." *The Journal of Economic History* 78 (4): 1179–1209.

Clay, K., J. Lewis and E. Severnini (2019). "What Explains Cross-City Variation in Mortality During the 1918 Influenza Pandemic? Evidence from 438 U.S. Cities." *Economics & Human Biology* 35: 42–50.

Correia, S., S. Luck and E. Verner (2022). "Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu." *The Journal of Economic History* 82 (4): 917–957.

Currie, J. and H. Schwandt (2013). "Within-Mother Analysis of Seasonal Patterns in Health at Birth." *Proceedings of the National Academy of Sciences* 110 (30): 12265–12270.

Eriksson, J. G., E. Kajantie, C. Osmond, K. Thornburg and D. J. P. Barker (2010). "Boys Live Dangerously in the Womb." *American Journal of Human Biology* 22 (3): 330–335.

Fletcher, J. M. (2018). "New Evidence on the Impacts of Early Exposure to the 1918 Influenza Pandemic on Old-Age Mortality." *Biodemography and Social Biology* 64 (2): 123–126.

Garrett, T. A. (2008). "Pandemic Economics: The 1918 Influenza and Its Modern-Day Implications." *Federal Reserve Bank of St. Louis Review* 90 (2): 75–93. Retrieved from https://doi.org/10.20955/r.90.74-94 in October 2025.

Goldstein, J. R., M. Alexander, C. Breen, A. M. González, F. Menares, M. Osborne, M. Snyder and U. Yildirim (2023). *CenSoc Mortality File: Version 3.0*. Berkeley: University of California.

Goldstein, J. R., M. Alexander, C. Breen, A. M. González, F. Menares, M. Osborne, M. Snyder and U. Yildirim (2025). *CenSoc Mortality File: Version 4.0*. Berkeley: University of California.

Halpern-Manners, A., J. Helgertz, J. R. Warren and E. Roberts (2020). "The Effects of Education on Mortality: Evidence from Linked U.S. Census and Administrative Mortality Data." *Demography* 57 (4): 1513–1541.

Hatchett, R. J., C. E. Mecher and M. Lipsitch (2007). "Public Health Interventions and Epidemic Intensity During the 1918 Influenza Pandemic." *Proceedings of the National Academy of Sciences* 104 (18): 7582–7587.

Helgertz, J. and T. Bengtsson (2019). "The Long-Lasting Influenza: The Impact of Fetal Stress During the 1918 Influenza Pandemic on Socioeconomic Attainment and Health in Sweden, 1968–2012." *Demography* 56 (4): 1389–1425.

Kennedy, E. H. (2023). "Towards Optimal Doubly Robust Estimation of Heterogeneous Causal Effects." *Electronic Journal of Statistics* 17 (2): 3008–3049.

Knaus, M. C. (2022). "Double Machine Learning-Based Programme Evaluation under Unconfoundedness." *The Econometrics Journal* 25 (3): 602–627.

Kraemer, S. (2000). "The Fragile Male." *BMJ: British Medical Journal* 321 (7276): 1609–1612.

Lee, D. S. (2009). "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies* 76 (3): 1071–1102.

Lin, M.-J. and E. M. Liu (2014). "Does In Utero Exposure to Illness Matter? The 1918 Influenza Epidemic in Taiwan as a Natural Experiment." *Journal of Health Economics* 37: 152–163.

Lutz, W., E. Striessnig, A. Dimitrova, S. Ghislandi, A. Lijadi, C. Reiter, S. Spitzer and D. Yildiz (2021). "Years of Good Life Is a Well-Being Indicator Designed to Serve Research on Sustainability." *Proceedings of the National Academy of Sciences* 118 (12): e1907351118.

Markel, H., H. B. Lipman, J. A. Navarro, A. Sloan, J. R. Michalsen, A. M. Stern and M. S. Cetron (2007). "Nonpharmaceutical Interventions Implemented by US Cities During the 1918–1919 Influenza Pandemic." *JAMA* 298 (6): 644–654.

Mazumder, B., D. Almond, K. Park, E. M. Crimmins and C. E. Finch (2010). "Lingering Prenatal Effects of the 1918 Influenza Pandemic on Cardiovascular Disease." *Journal of Developmental Origins of Health and Disease* 1 (1): 26–34.

Myrskylä, M. and A. Fenelon (2012). "Maternal Age and Offspring Adult Health: Evidence from the Health and Retirement Study." *Demography* 49 (4): 1231–1257.

Myrskylä, M., N. K. Mehta and V. W. Chang (2013). "Early Life Exposure to the 1918 Influenza Pandemic and Old-Age Mortality by Cause of Death." *American Journal of Public Health* 103 (7): e83–e90.

Neelsen, S. and T. Stratmann (2012). "Long-Run Effects of Fetal Influenza Exposure: Evidence from Switzerland." *Social Science & Medicine* 74 (1): 58–66.

Nelson, R. E. (2010). "Testing the Fetal Origins Hypothesis in a Developing Country: Evidence from the 1918 Influenza Pandemic." *Health Economics* 19 (10): 1181–1192.

Ogasawara, K. (2022). "Pandemic Influenza and Gender Imbalance: Mortality Selection Before Births." *Social Science & Medicine* 311: 115299.

Oster, E. (2019). "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2): 187–204.

Ruggles, S., S. Flood, M. Sobek, D. Backman, G. Cooper, J. A. Rivera Drew, S. Richards, R. Rodgers, J. Schroeder and K. C. W. Williams (2025). *IPUMS USA: Version 16.0* [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V16.0 (accessed in September 2025).

Ruggles, S., M. A. Nelson, M. Sobek, C. A. Fitch, R. Goeken, J. D. Hacker, E. Roberts and J. R. Warren (2024). *IPUMS Ancestry Full Count Data: Version 4.0* [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D014.V4.0 (accessed in September 2025).

Ruggles, S., N. Ozder, C. A. Fitch, M. Sobek, J. A. Rivera Drew, J. D. Hacker, J. Helgertz, C. Lonobile, M. A. Nelson, E. Roberts and J. R. Warren (2025). *IPUMS Multigenerational Longitudinal Panel: Version 2.0* [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D016.V2.0 (accessed in September 2025).

Semenova, V. and V. Chernozhukov (2021). "Debiased Machine Learning of Conditional Average Treatment Effects and Other Causal Functions." *The Econometrics Journal* 24 (2): 264–289.

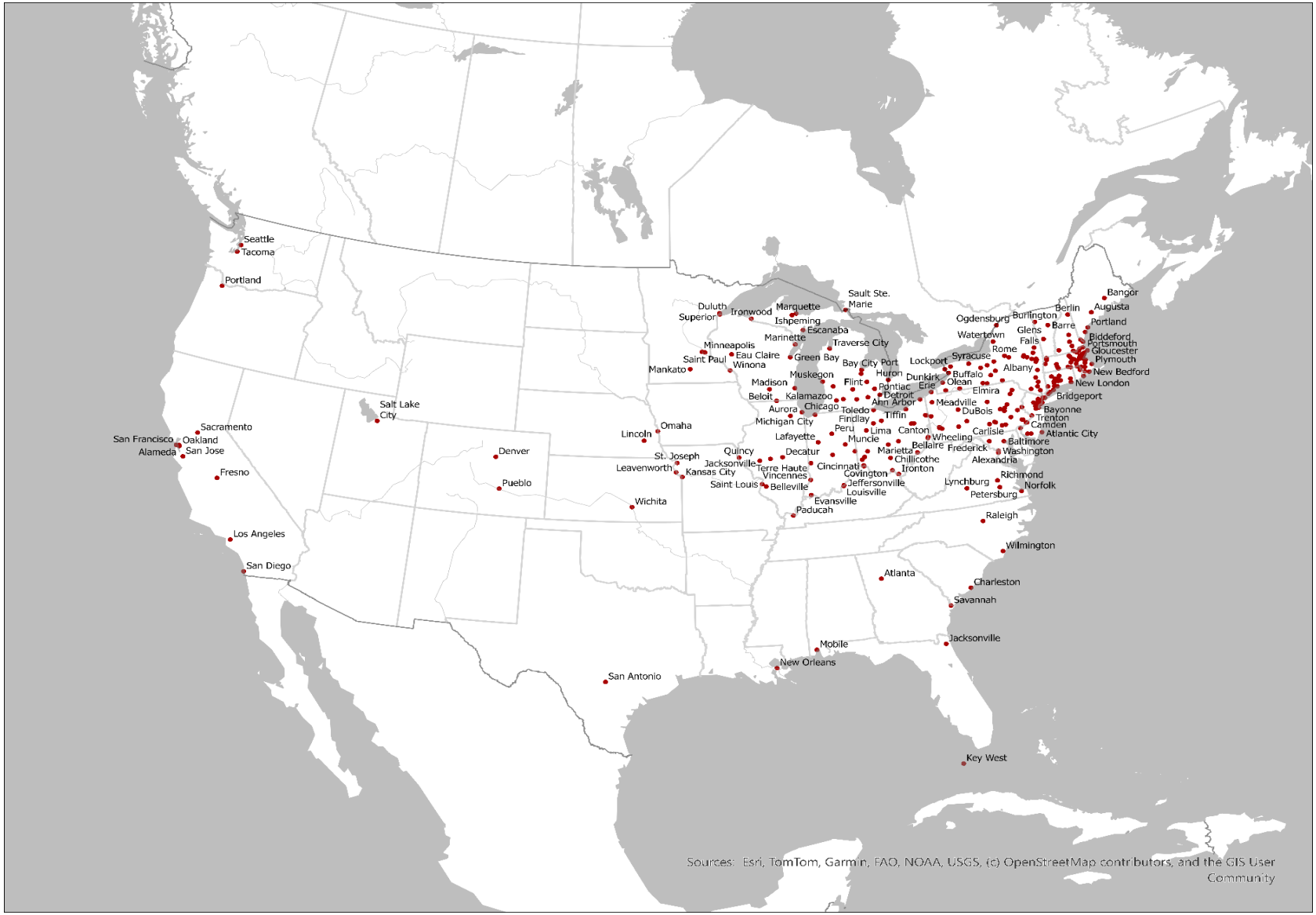


Figure 1: Cities in the Estimation Sample

Notes: The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. Individuals born in 1918 and 1919 are included in the estimation sample. The estimation sample includes no cities from Alaska or Hawaii.

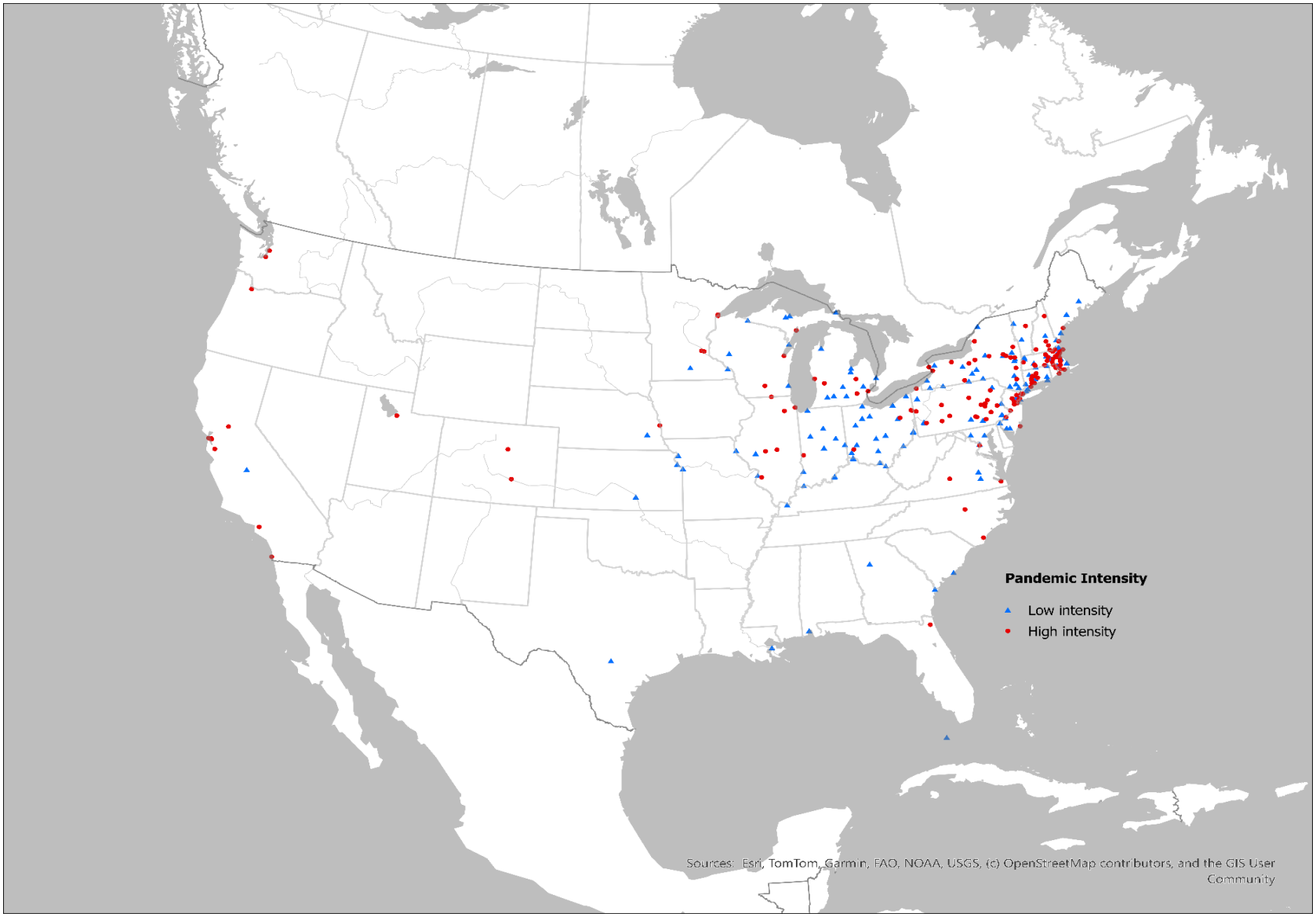


Figure 2: Cities in the Estimation Sample by Pandemic Intensity Level

Notes: The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. The median of the standardized city-level excess influenza mortality measure is used to define low-intensity cities (less than or equal to the median) and high-intensity cities (greater than the median). Individuals born in 1918 and 1919 are included in the estimation sample. The estimation sample includes no cities from Alaska or Hawaii.

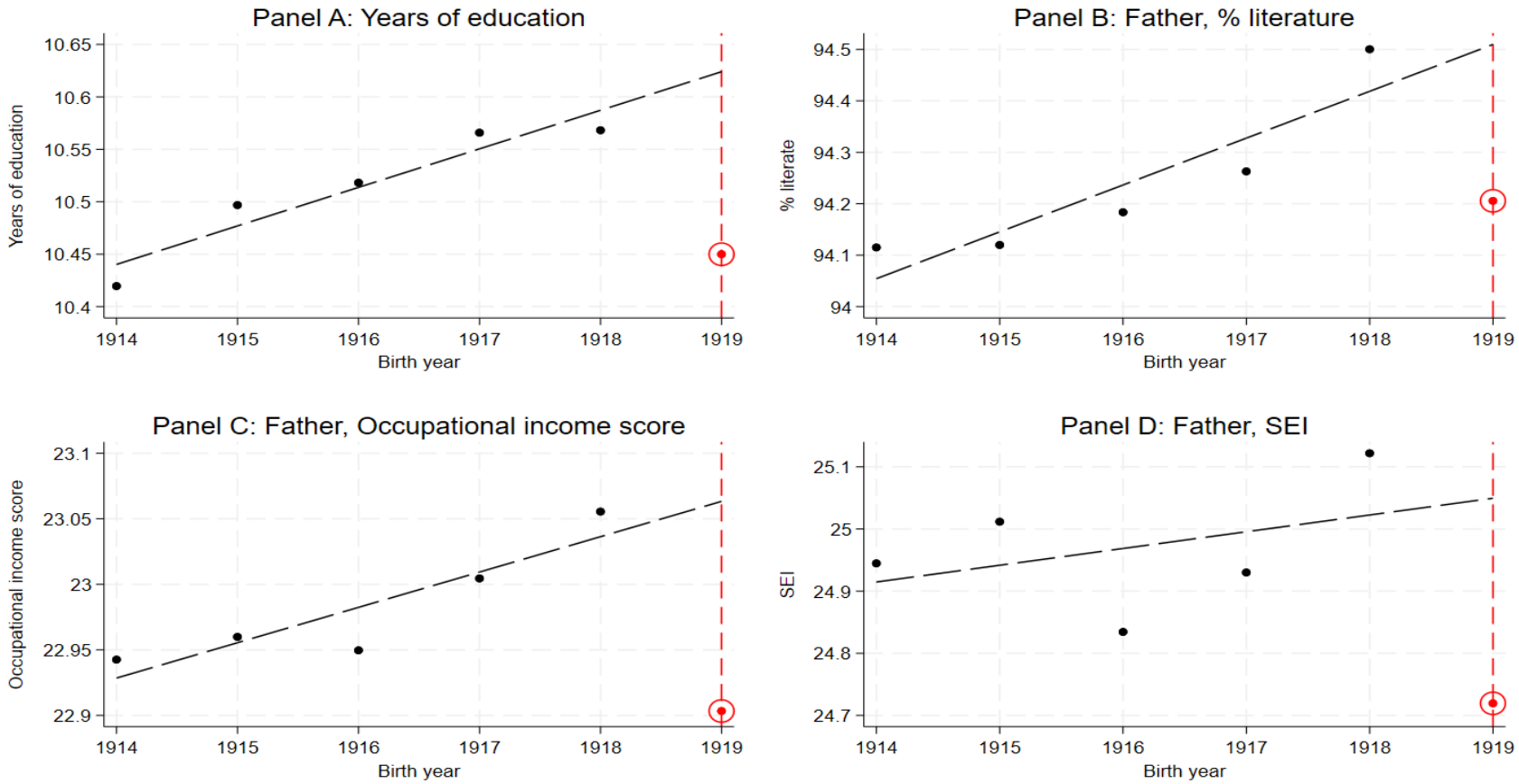


Figure 3: Own Years of Education and Paternal Characteristics by Birth Cohort

Notes: The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. Individuals born between 1914 and 1919 are included in the estimation sample. SEI stands for the Duncan Socioeconomic Index provided in the census data. Dashed lines in all panels depict OLS estimated trends that are linear in birth year.

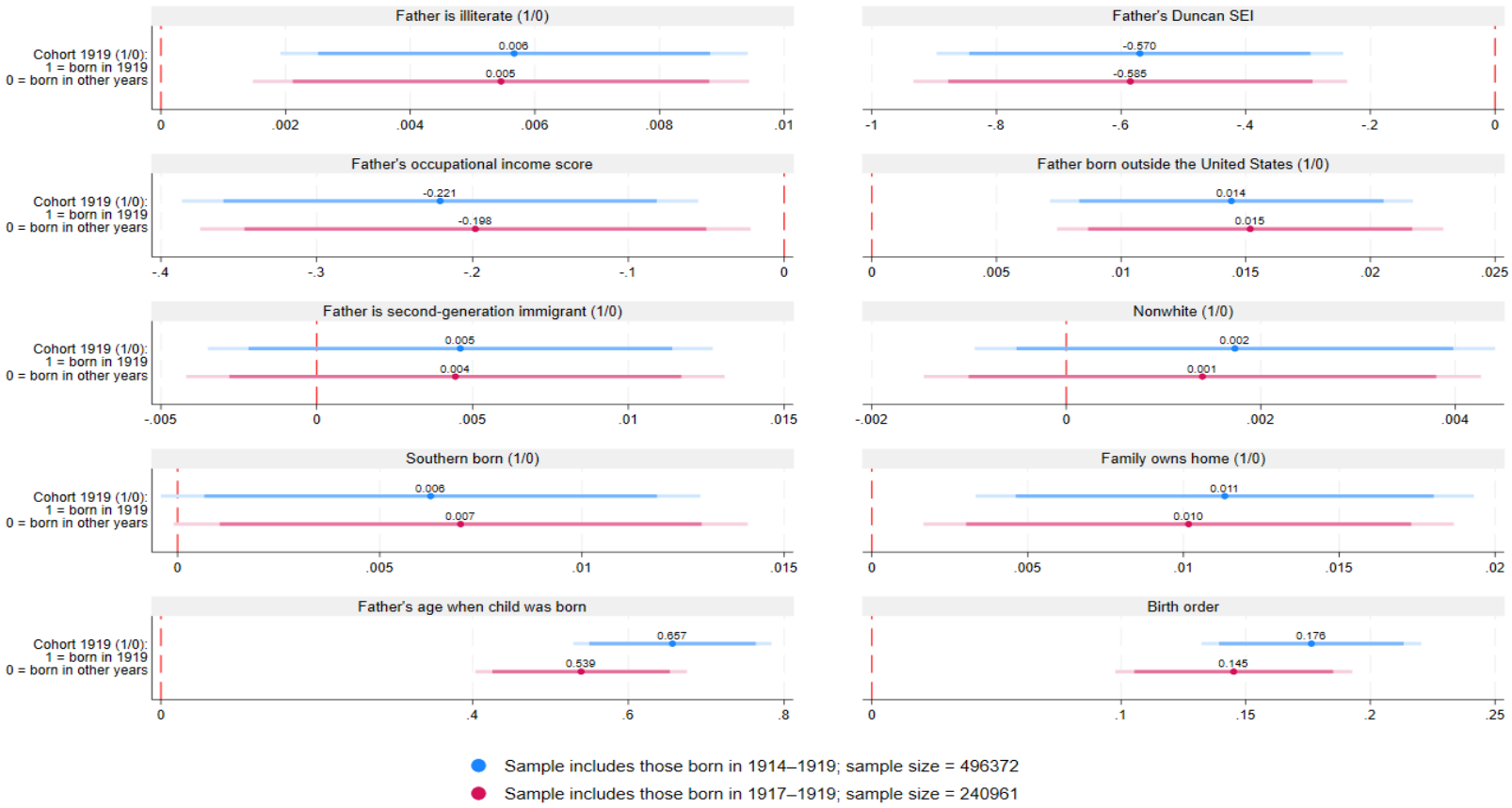


Figure 4: Estimated Deviation of 1919 Birth Cohort's Background Characteristics Relative to Other Birth Cohorts

Notes: The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. Each panel reports results from a separate OLS regression. The dependent variable is described in the panel title. Each regression model includes a birth cohort indicator (1/0) and controls for year of birth (linear and quadratic terms). Robust (to heteroskedasticity) standard errors are used to construct the confidence intervals. The figure reports point estimates and confidence intervals at the 95% and 90% levels, with the 90% intervals shown in a darker color.

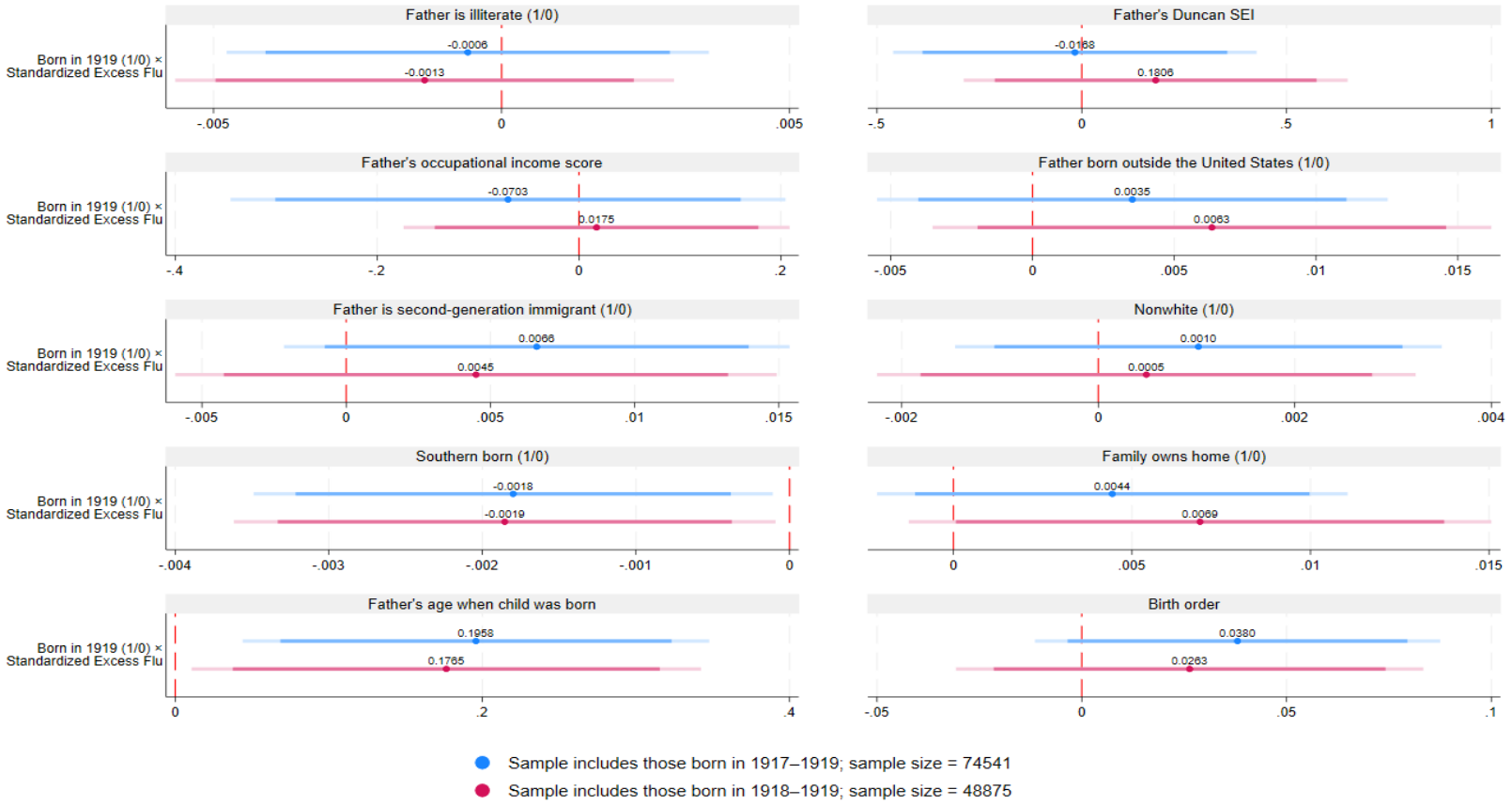


Figure 5: Assessing Whether Background Differences Are Related to Pandemic Intensity

Notes: The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. Each panel reports results from a separate OLS regression. The dependent variable is described in the panel title. Each regression model includes a birth cohort indicator (1/0) multiplied by the standardized city-level excess influenza mortality measure. This standardized measure is defined as city-level excess influenza deaths divided by the sample mean. Each regression model also includes birth year fixed effects and residential city in 1920 fixed effects. Standard errors are clustered at the residential city in 1920 level. The figure reports point estimates and confidence intervals at the 95% and 90% levels, with the 90% intervals shown in a darker color.

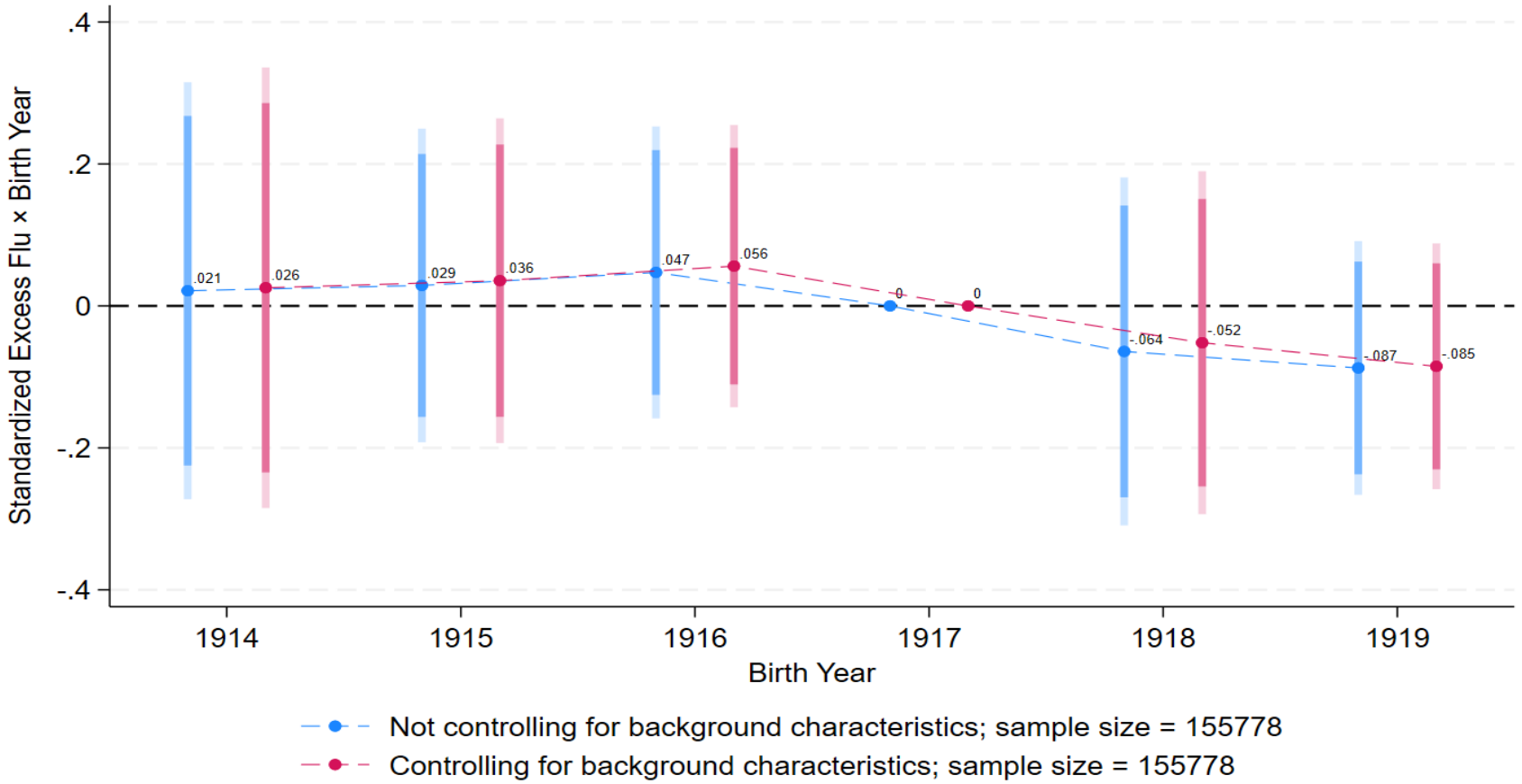


Figure 6: Event-Study Estimates of the Impact of Pandemic Exposure on Age at Death

Notes: The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. Individuals born between 1914 and 1919 are included in the estimation sample. The dependent variable is age at death (measured in years). Each OLS regression model includes birth cohort indicators (excluding the birth cohort of year 1917, which is used as the reference group), each of which is multiplied by the standardized city-level excess influenza mortality measure. This standardized measure is defined as city-level excess influenza deaths divided by the sample mean. All regression models control for birth year fixed effects and residential city in 1920 fixed effects. In regression models that also control from background characteristics, these background characteristics are from the 1920 census, including: dummy variables for own race, own state of birth, father's state of birth (if within the U.S.) or country of birth (if outside the U.S.) of birth, mother's state of birth (if within the U.S.) or country of birth (if outside the U.S.), father's occupation code (based on the Census Bureau occupational classification system), a second-generation immigrant father, whether the father moved outside of birthplace before child was born, an illiterate father, an illiterate mother, whether the mother was in the labor force, whether the family owned the home; father's age when child was born; mother's age when child was born; the difference between father's and mother's ages; and birth order. Standard errors are clustered at the residential city in 1920 level. The figure reports point estimates and confidence intervals at the 95% and 90% levels, with the 90% intervals shown in a darker color.

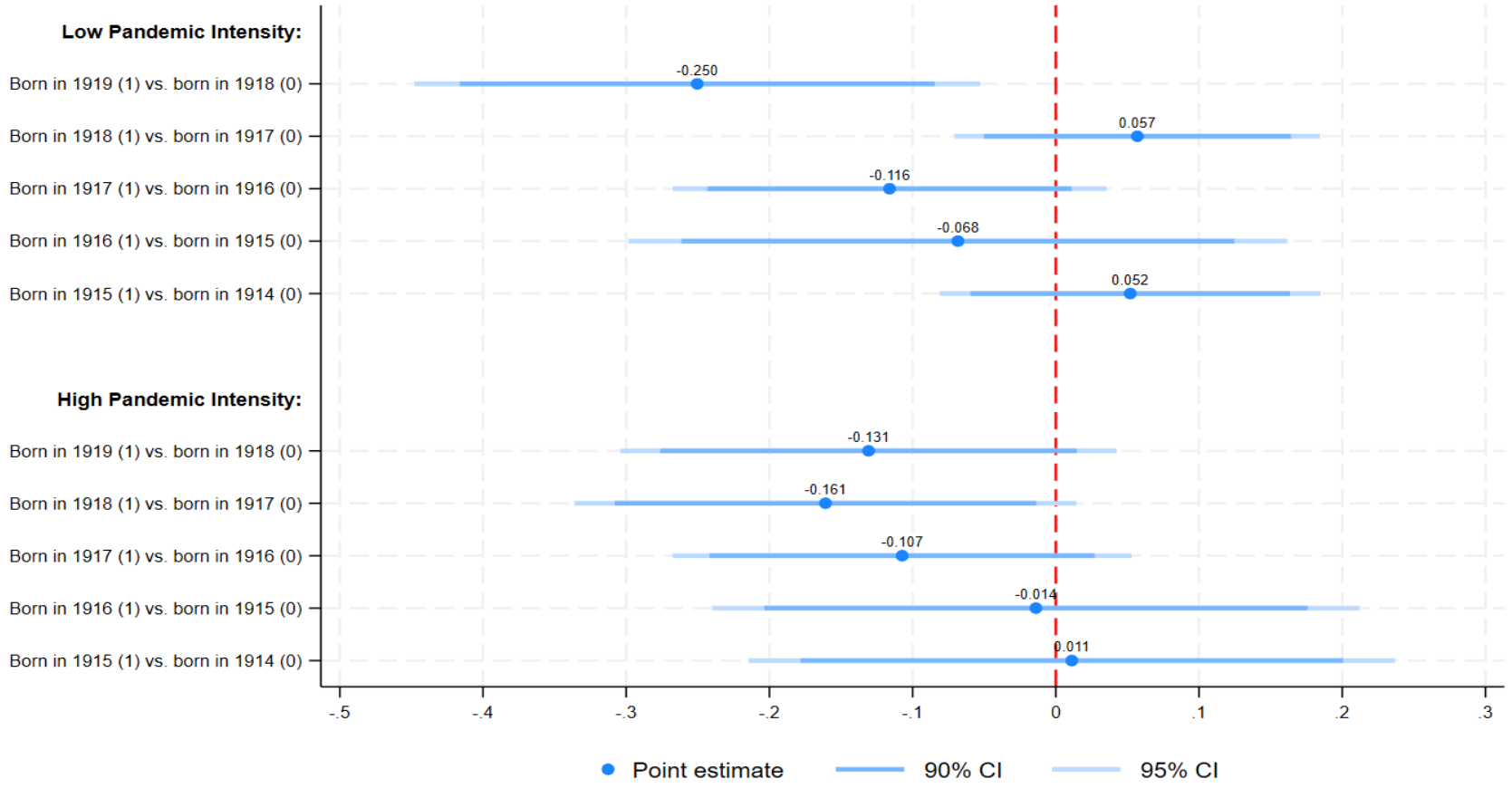


Figure 7: Differences in Age at Death of Males Born in Adjacent Years

Notes: The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. The median of the standardized city-level excess influenza mortality measure is used to define low-intensity cities (less than or equal to the median) and high-intensity cities (greater than the median). The figure reports point estimates and confidence intervals (CI) of the differences in age at death (measured in years) between adjacent birth cohorts. These differences are estimated by OLS using regression models that control for birth month fixed effects and residential city in 1920 fixed effects, as well as background characteristics. These background characteristics are from the 1920 census, including: dummy variables for own race, own state of birth, father's state of birth (if within the U.S.) or country of birth (if outside the U.S.) of birth, mother's state of birth (if within the U.S.) or country of birth (if outside the U.S.), father's occupation code (based on the Census Bureau occupational classification system), a second-generation immigrant father, whether the father moved outside of birthplace before child was born, an illiterate father, an illiterate mother, whether the mother was in the labor force, whether the family owned the home; father's age when child was born; mother's age when child was born; the difference between father's and mother's ages; and birth order. Standard errors are clustered at the residential city in 1920 level.

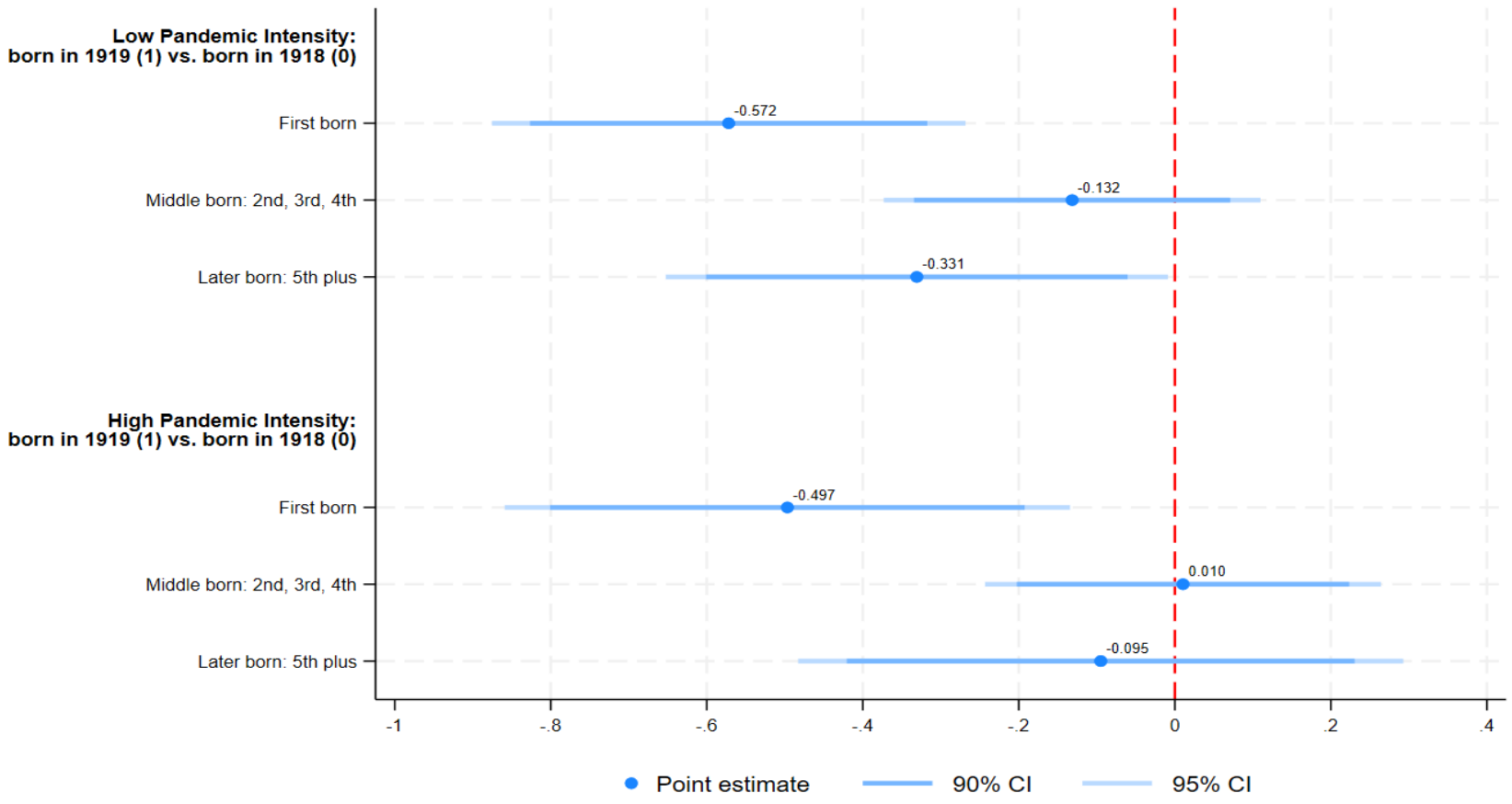


Figure 8: Differences in Age at Death of Males Born in 1919 Relative to Those Born in 1918 by Birth Order

Notes: the same as the notes of Figure 7.

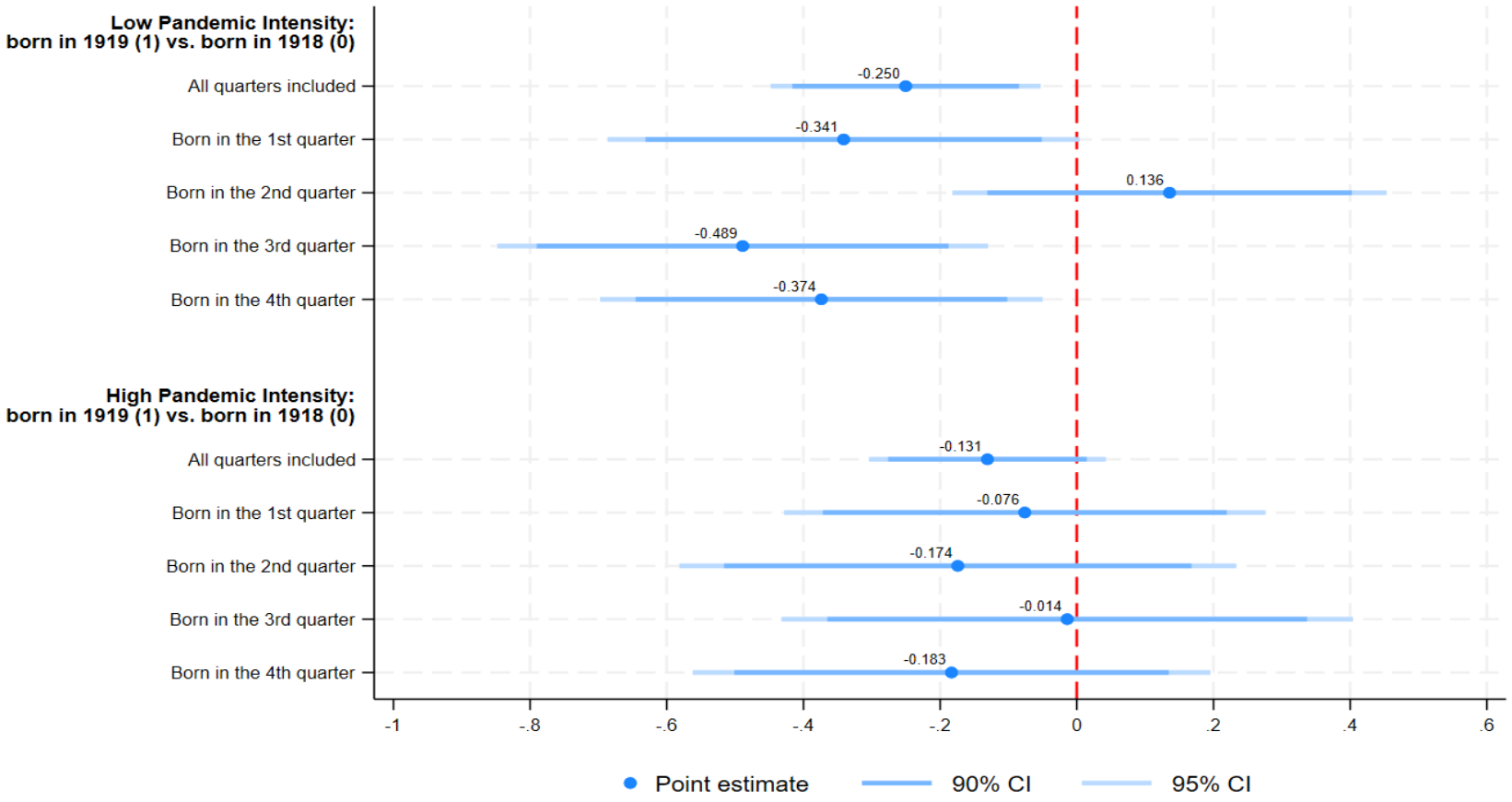


Figure 9: Differences in Age at Death of Males Born in 1919 Relative to Those Born in 1918 by Quarter of Birth

Notes: the same as the notes of Figure 7.

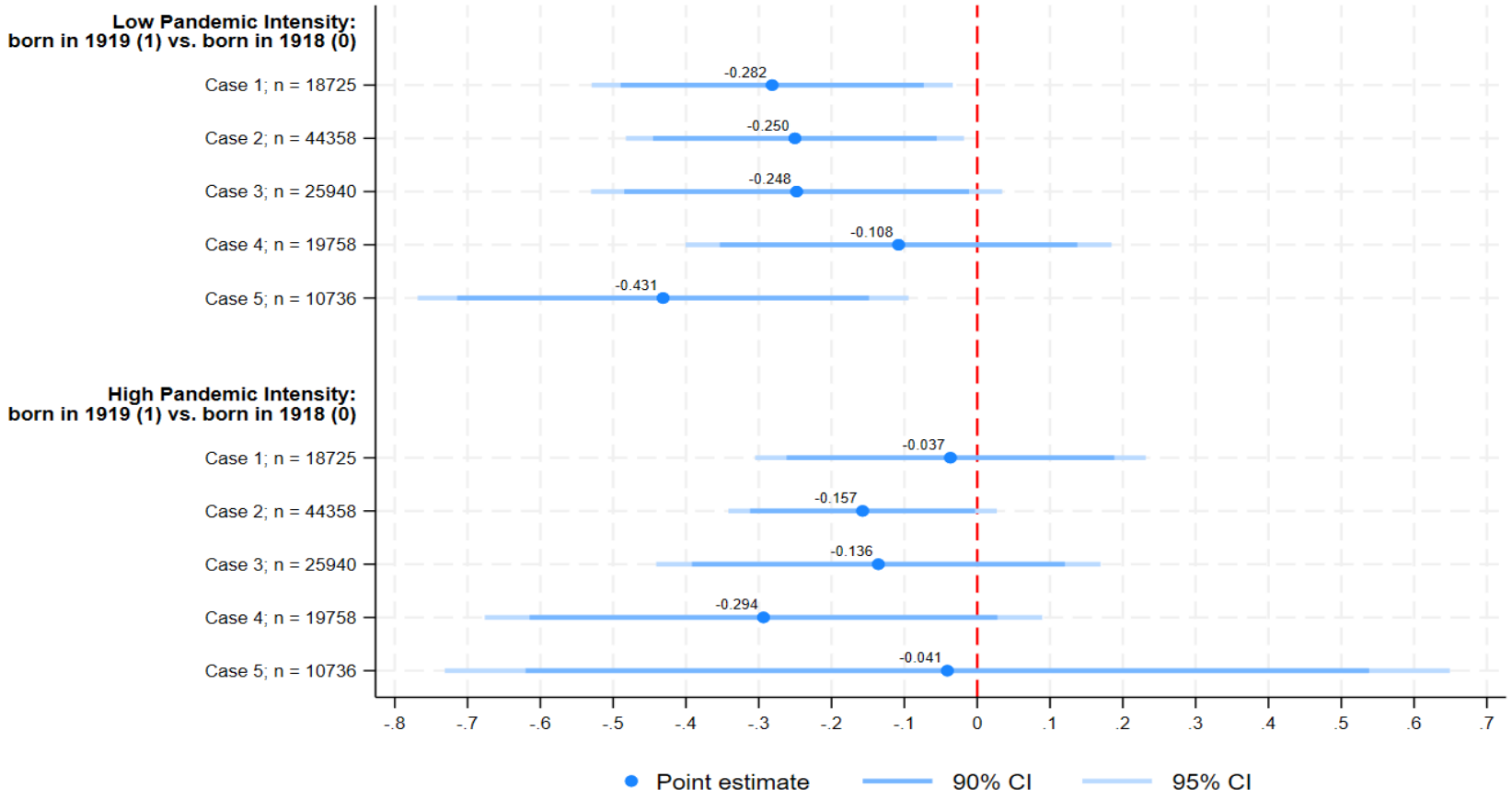


Figure 10: Differences in Age at Death of Males Born in 1919 Relative to Those Born in 1918, Estimated Using Alternative Specifications

Notes: the same as the notes of Figure 7 except that:

In Case 1, low-intensity cities in the estimation sample are cities in Michigan, Minesota and Wisconsin (the bottom three states based on the 1918 influenza mortality rate rank in table 1 of Garrett (2008); high-intensity cities in the estimation sample are cities in Pennsylvania, Maryland and New Jersey (the top three states based on the 1918 influenza mortality rate rank in table 1 of table 1 of Garrett (2008).

In Case 2, estimation uses a subsample including states that have both low-intensity cities and high-intensity cities (explained in the notes section of Figure 2).

In Case 3, estimation uses a subsample including states that have both low-intensity cities and high-intensity cities (explained in the notes section of Figure 2), and also states in the northeast region of the United States (i.e., PA, NY, NJ, CT, RI, MA, NH, VT and ME).

In Case 4, estimation uses a subsample including those living in rural areas in 1920 of the six states used in Case 1; residential city in 1920 fixed effects are replaced by residential county in 1920 fixed effects; standard errors are clustered at the residential county in 1920 level.

In Case 5, estimation uses Philadelphia, PA (the residential city in 1920) as the high-intensity city and New York City, New York (the residential city in 1920) as the low-intensity city; residential city in 1920 fixed effects are no longer used; robust (to heteroskedasticity) standard errors are used to construct the confidence intervals (CI).

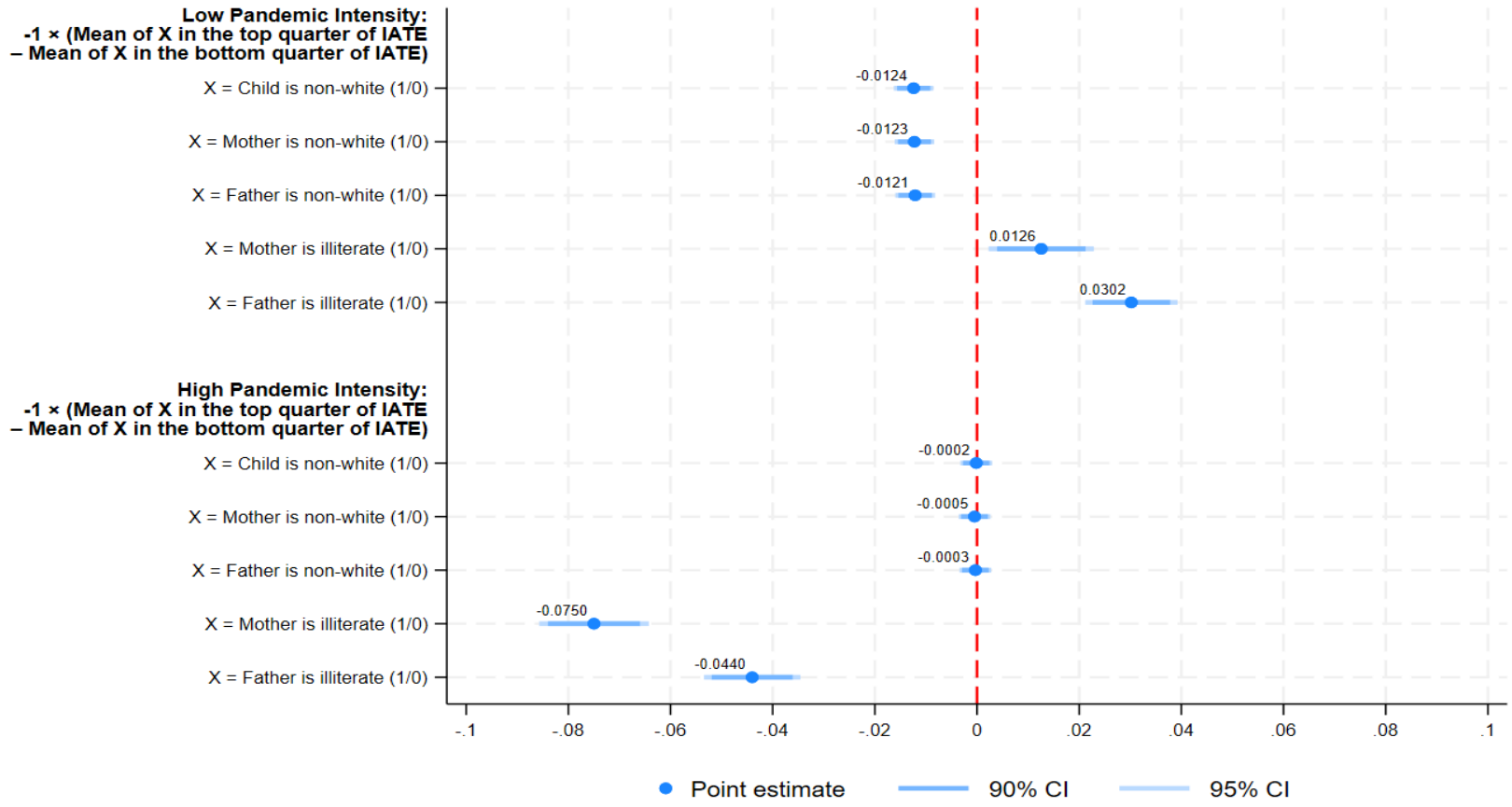


Figure 11: Classification Analysis of the Groups Constructed Based on the Sorted IATE

Notes: IATE stands for individualized average treatment effects. The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. The median of the standardized city-level excess influenza mortality measure is used to define low-intensity cities (less than or equal to the median) and high-intensity cities (greater than the median). Individuals born in 1918 and 1919 are included in the estimation sample. Generalized random forest and double/debiased machine learning (using 10-fold cross-fitting) are used to estimate IATE. In each estimation, the dependent variable is age at death (measured in years); the treatment group consists of those born in 1919; the control group consists of those born in 1918; control variables include birth month dummy variables, residential city in 1920 dummy variables, and the background characteristics listed in Figure 7. Estimations are done separately for low-intensity cities and high-intensity cities. At the end of each estimation, the sample is divided into four groups based on quartiles of the IATE estimates. The sorting is based on the IATE estimated values as they are (not absolute values). We multiply the (mean of X in the top quarter of IATE – mean of X in the bottom quarter of IATE) by -1, given that more negative IATE values indicate greater effects on longevity. The figure reports point estimates and confidence intervals (CI) of the differences in the means of demographic variables between the top quarter and the bottom quarter.

Table 1: Summary Statistics

	Low pandemic intensity	High pandemic intensity	Total
Panel A: Variables obtained from CenSoc-DMF, Beach et al. (2022) and the 1920 full-count census			
Number of observations	26,759 (52.3%)	24,381 (47.7%)	51,140 (100.0%)
Age at Death (Years)	78.612 (7.692)	78.635 (7.786)	78.623 (7.737)
Birth cohort (1/0): 1 for 1919 and 0 for 1918	0.477 (0.499)	0.474 (0.499)	0.475 (0.499)
City-level excess 1918 influenza deaths (Beach et al. 2022)	22.158 (6.027)	52.325 (28.025)	36.540 (24.909)
Non-white (1/0)	0.017 (0.129)	0.010 (0.101)	0.014 (0.116)
Birth order	3.617 (2.511)	3.295 (2.320)	3.464 (2.427)
Second-generation immigrant father (1/0)	0.687 (0.464)	0.748 (0.434)	0.716 (0.451)
Father moved outside of birthplace before child was born (1/0)	0.578 (0.494)	0.662 (0.473)	0.618 (0.486)
Illiterate father (1/0)	0.060 (0.238)	0.068 (0.252)	0.064 (0.244)
Illiterate mother (1/0)	0.082 (0.275)	0.091 (0.288)	0.087 (0.281)
Mother was in the labor force (1/0)	0.030 (0.172)	0.032 (0.175)	0.031 (0.173)
Family owned the home (1/0)	0.250 (0.433)	0.292 (0.455)	0.270 (0.444)
Father's age when child was born	32.074 (7.060)	32.387 (7.132)	32.224 (7.096)
Mother's age when child was born	28.040 (5.842)	28.250 (5.805)	28.140 (5.825)
Difference between father's and mother's ages	4.034 (4.953)	4.138 (5.224)	4.084 (5.084)
Panel B: Variable obtained from the 1920–1940 linked census			
Number of observations	26,261 (52.2%)	24,064 (47.8%)	50,325 (100.0%)
Years of education in 1940	11.211 (2.399)	11.230 (2.227)	11.220 (2.318)
Panel C: Variable obtained from the 1920–1940 linked census			
Number of observations	22,469 (52.4%)	20,446 (47.6%)	42,915 (100.0%)
Employed in 1940 (1/0)	0.800 (0.400)	0.802 (0.398)	0.801 (0.399)
Panel D: Variable obtained from the 1920–1940 linked census			
Number of observations	20,421 (52.3%)	18,651 (47.7%)	39,072 (100.0%)
Duncan Socioeconomic Index in 1940	30.069 (17.941)	27.870 (17.448)	29.019 (17.741)
Panel E: Variable obtained from the the 1920–1940 linked census			
Number of observations	3,830 (51.6%)	3,595 (48.4%)	7,425 (100.0%)
Wage and salary income in 1940 (in 1999 dollars)	6,011.029 (5,592.295)	6,040.901 (5,760.620)	6,025.492 (5,674.054)
Panel F: Variable obtained from the 1940–1950 linked census			
Number of observations	3,830 (51.6%)	3,595 (48.4%)	7,425 (100.0%)
Wage and salary income in 1950 (in 1999 dollars)	17,494.883 (12,089.428)	17,906.273 (12,002.919)	17,694.067 (12,048.564)

Notes: The mean and standard deviation (in parenthesis) are reported in the table. These summary statistics are based on the estimation sample that uses the following data: CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022). The median of the standardized city-level excess influenza mortality measure is used to define low-intensity cities (less than or equal to the median) and high-intensity cities (greater than the median). Individuals born in 1918 and 1919 are included in the estimation sample.

Table 2: Differences in Age at Death of Males Born in Adjacent Years in Low Pandemic Intensity Cities

	(1)	(2)	(3)	(4)
Panel A				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.264 ***	-0.246 **	-0.252 **	-0.250 **
(Standard error)	(0.097)	(0.102)	(0.102)	(0.101)
[P-value]	[0.007]	[0.017]	[0.014]	[0.014]
Panel B				
Birth cohort (1/0): 1 for 1918 and 0 for 1917	0.030	0.053	0.055	0.057
(Standard error)	(0.067)	(0.066)	(0.065)	(0.065)
[P-value]	[0.660]	[0.417]	[0.398]	[0.383]
Panel C				
Birth cohort (1/0): 1 for 1917 and 0 for 1916	-0.112	-0.114	-0.116	-0.116
(Standard error)	(0.081)	(0.078)	(0.077)	(0.077)
[P-value]	[0.166]	[0.143]	[0.135]	[0.134]
Panel D				
Birth cohort (1/0): 1 for 1916 and 0 for 1915	-0.090	-0.052	-0.068	-0.068
(Standard error)	(0.121)	(0.121)	(0.117)	(0.117)
[P-value]	[0.457]	[0.669]	[0.564]	[0.561]
Panel E				
Birth cohort (1/0): 1 for 1915 and 0 for 1914	0.054	0.043	0.051	0.052
(Standard error)	(0.068)	(0.068)	(0.068)	(0.068)
[P-value]	[0.424]	[0.533]	[0.456]	[0.444]
Control variables used for Panels A through E				
Background characteristics	No	Yes	Yes	Yes
City FE	No	No	Yes	Yes
Birth month FE	No	No	No	Yes

Notes: Sample sizes in Panels A, B, C, D and E respectively are: 51,140; 53,696; 53,137; 52,087; 51,501. The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. The median of the standardized city-level excess influenza mortality measure is used to define low-intensity cities (less than or equal to the median) and high-intensity cities (greater than the median). Each panel reports a separate set of OLS estimation results. Differences in age at death (measured in years) between adjacent birth cohorts are estimated by OLS, using regression models that include (or exclude) birth month fixed effects (FE) and residential city in 1920 fixed effects (city FE), as well as background characteristics. These background characteristics are from the 1920 census, including: dummy variables for own race, own state of birth, father's state of birth (if within the U.S.) or country of birth (if outside the U.S.) of birth, mother's state of birth (if within the U.S.) or country of birth (if outside the U.S.), father's occupation code (based on the Census Bureau occupational classification system), a second-generation immigrant father, whether the father moved outside of birthplace before child was born, an illiterate father, an illiterate mother, whether the mother was in the labor force, whether the family owned the home; father's age when child was born; mother's age when child was born; the difference between father's and mother's ages; and birth order. Standard errors are clustered at the residential city in 1920 level. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

Table 3: Differences in Age at Death of Males Born in Adjacent Years in High Pandemic Intensity Cities

	(1)	(2)	(3)	(4)
Panel A				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.134	-0.136	-0.131	-0.131
(Standard error)	(0.087)	(0.088)	(0.088)	(0.088)
[P-value]	[0.122]	[0.125]	[0.136]	[0.140]
Panel B				
Birth cohort (1/0): 1 for 1918 and 0 for 1917	-0.196 **	-0.160 *	-0.164 *	-0.161 *
(Standard error)	(0.082)	(0.087)	(0.089)	(0.089)
[P-value]	[0.018]	[0.069]	[0.067]	[0.073]
Panel C				
Birth cohort (1/0): 1 for 1917 and 0 for 1916	-0.084	-0.100	-0.106	-0.107
(Standard error)	(0.086)	(0.083)	(0.082)	(0.082)
[P-value]	[0.328]	[0.232]	[0.198]	[0.190]
Panel D				
Birth cohort (1/0): 1 for 1916 and 0 for 1915	-0.037	-0.023	-0.017	-0.014
(Standard error)	(0.111)	(0.114)	(0.115)	(0.115)
[P-value]	[0.737]	[0.842]	[0.885]	[0.904]
Panel E				
Birth cohort (1/0): 1 for 1915 and 0 for 1914	-0.028	0.005	0.009	0.011
(Standard error)	(0.109)	(0.115)	(0.115)	(0.115)
[P-value]	[0.799]	[0.968]	[0.939]	[0.923]
Control variables used for Panels A through E				
Background characteristics	No	Yes	Yes	Yes
City FE	No	No	Yes	Yes
Birth month FE	No	No	No	Yes

Notes: the same as the notes of Table 2. * p -value < 0.1; ** p -value < 0.05; *** p -value < 0.01.

Table 4: Effects of In Utero Exposure to the 1918 Influenza Pandemic on Other Outcomes for Males in Low or High Pandemic Intensity Cities

	(1)	(2)	(3)	(4)	(5)	(6)
Pandemic Intensity:	Low	Low	Low	High	High	High
Panel A: Years of education in 1940						
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.054 *	-0.053	-0.051	-0.090 **	-0.087 **	-0.088 **
(Standard error)	(0.033)	(0.032)	(0.036)	(0.039)	(0.039)	(0.038)
[P-value]	[0.098]	[0.101]	[0.155]	[0.021]	[0.026]	[0.022]
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Birth month FE	Yes	Yes	Yes	Yes	Yes	Yes
1920 Residence city FE	No	Yes	Yes	No	Yes	Yes
1940 Residence county FE	No	No	Yes	No	No	Yes
Panel B: Employed in 1940 (1/0)						
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.038 ***	-0.037 ***	-0.037 ***	-0.042 ***	-0.041 ***	-0.042 ***
(Standard error)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
[P-value]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Birth month FE	Yes	Yes	Yes	Yes	Yes	Yes
1920 Residence city FE	No	Yes	Yes	No	Yes	Yes
1940 Residence county FE	No	No	Yes	No	No	Yes
Panel C: Duncan Socioeconomic Index in 1940						
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-1.106 ***	-1.157 ***	-1.122 ***	-0.939 ***	-0.891 ***	-0.884 ***
(Standard error)	(0.194)	(0.192)	(0.197)	(0.303)	(0.308)	(0.306)
[P-value]	[0.000]	[0.000]	[0.000]	[0.002]	[0.004]	[0.004]
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Birth month FE	Yes	Yes	Yes	Yes	Yes	Yes
1920 Residence city FE	No	Yes	Yes	No	Yes	Yes
1940 Residence county FE	No	No	Yes	No	No	Yes
Panel D: Wage and salary income in 1940 (in 1999 dollars)						
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-1532.165 ***	-1562.658 ***	-1641.279 ***	-1187.626 ***	-1196.806 ***	-1153.710 ***
(Standard error)	(196.347)	(205.553)	(215.698)	(189.290)	(185.146)	(190.837)
[P-value]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Birth month FE	Yes	Yes	Yes	Yes	Yes	Yes
1920 Residence city FE	No	Yes	Yes	No	Yes	Yes
1940 Residence county FE	No	No	Yes	No	No	Yes
Panel E: Wage and salary income in 1950 (in 1999 dollars)						
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-564.977 **	-635.814 **	-635.760 **	854.508 **	797.343 **	672.179 *
(Standard error)	(260.535)	(285.931)	(291.543)	(342.150)	(376.012)	(378.723)
[P-value]	[0.031]	[0.027]	[0.030]	[0.013]	[0.035]	[0.077]
Background characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Birth month FE	Yes	Yes	Yes	Yes	Yes	Yes
1920 Residence city FE	No	Yes	Yes	No	Yes	Yes
1950 Residence county FE	No	No	Yes	No	No	Yes

Notes: Sample sizes in Panels A, B, C, D and E respectively are: 50,325; 42,915; 39,072; 7,425; 7,425. The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. The median of the standardized city-level excess influenza mortality measure is used to define low-intensity cities (less than or equal to the median) and high-intensity cities (greater than the median). Each panel reports a separate set of OLS estimation results. Differences in X, where X is the dependent variable described in the panel title, between adjacent birth cohorts are estimated by OLS, using regression models that control for fixed effects (FE) and background characteristics. These background characteristics are from the 1920 census, including: dummy variables for own race, own state of birth, father's state of birth (if within the U.S.) or country of birth (if outside the U.S.) of birth, mother's state of birth (if within the U.S.) or country of birth (if outside the U.S.), father's occupation code (based on the Census Bureau occupational classification system), a second-generation immigrant father, whether the father moved outside of birthplace before child was born, an illiterate father, an illiterate mother, whether the mother was in the labor force, whether the family owned the home; father's age when child was born; mother's age when child was born; the difference between father's and mother's ages; and birth order. Standard errors are clustered at the residential city in 1920 level. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

Appendices for "When Less Still Harmed: Old-Age Mortality Effects of In Utero Exposure to the 1918 Influenza Pandemic for Males in Low-Intensity Cities"[1]

By MUZHE YANG[2]

March 2026

Appendix A: Robustness Checks

Alternative Specifications.—Appendix Table A1 and Appendix Table A2 report the estimates of the in utero pandemic exposure on old-age mortality for low- and high-intensity cities (focusing on the 1919 and 1918 birth cohorts), respectively. Panel A of both tables reports the results of estimations that use the sample mean of the standardized city-level excess influenza mortality measure (explained in the main text of the paper) rather than the median. Panel B of both tables reports the results of estimations that cluster standard errors at the county level, instead of the residential city in 1920 level. Panel C of both tables reports the results of using the absolute median value of the standardized city-level excess influenza mortality measure computed from the original data provided by Beach et al. (2022), rather than the median value computed from our study's estimation sample.

The main results on the effects of pandemic exposure on old-age mortality (discussed in Figure 7) are robust to these alternative specifications. Specifically, point estimates are very similar in the case examined in Panel A; statistical significance remain the same in the case examined in Panel B; estimation and inference results remain virtually identical (up to three decimal places) in the case examined in Panel C.

Sample Selection.—To address concerns about sample selection, we apply the bounding method proposed by Lee (2009), following Lin and Liu (2014) whose study is comparable to ours. In both studies, the treatment group (i.e., the 1919 birth cohort) is smaller than the control group (i.e., the

[1] The data used in this study were drawn from the Social Security Administration mortality records linked to the full-count 1940 U.S. Census through CenSoc, available at https://censoc.berkeley.edu/data/ (accessed in September 2025), and from U.S. Census data provided by IPUMS, available at https://usa.ipums.org/usa/ (accessed in September 2025). Under the data use agreements of both CenSoc and IPUMS, the author is not permitted to redistribute the downloaded data files. The paper provides the relevant source information and citations for these datasets to allow replication.

[2] Corresponding Author, Department of Economics, College of Business, Lehigh University, Bethlehem, PA 18015, muzheyang@lehigh.edu. Declarations of relevant or material financial interests that relate to the research described in this paper: none. Declarations of conflicts of interest: none.

1918 birth cohort), indicating excess attrition in the treatment group. Our analysis adopts the same assumption as Lin and Liu (2014): among those born in 1919, individuals who did not survive to the 1940 Census and to age 65 were the least healthy; under this assumption and using Lee's (2009) method, we trim 6.0 percent and 7.4 percent of the control group from below (i.e., lowest ages at death), using the 1920 full count census data and the 1920–1940 linked census data to determine attrition, respectively. Appendix Table A3 reports the estimates of the old-age mortality impact of pandemic exposure. Consistent with Lin and Liu (2014), the estimated effect becomes larger in magnitude. This pattern suggests that excess attribution in the treatment group, potentially driven by poorer underlying health, could lead the original estimates (reported in column 1 of Appendix Table A3) to understate the true impact of pandemic exposure.

Sensitivity Analysis.—We conduct sensitivity checks using the method proposed by Oster (2019) to assess robustness to selection on unobservables. Specifically, we evaluate the robustness of our estimated treatment effect—a reduction in age at death of 0.2504 years (significant at the 5% level) for the treatment group (i.e., the 1919 birth cohort) compared with the control group (i.e., the 1918 birth cohort) in low-intensity cities (reported in column 4 of Table 2). We examine how robust this estimate can be in the presence of selection on unobservables, given that the background characteristics and birth month fixed effects are controlled for, in addition to the use of birth year fixed effects and residential city in 1920 fixed effects in the baseline regression model. Results of this sensitivity analysis are reported in Appendix Figure A1. In both panels the x-axis represents what Oster (2019) refers to as the maximal R-squared. It is the R-squared obtained from a hypothetical regression model that includes all factors affecting the outcome variable (which is age at death in our case). These factors include those pre-determined (i.e., determined prior to the occurrence of the treatment) and those affected by the treatment. A key insight of Oster's method of linking selection bias due to unobservables with the stability of the treatment-effect estimates is the consideration of this maximal R-squared. In theory, the maximal R-squared could equal one if all contributing factors mentioned above are included in the regression model and the outcome variable is also measured without any error. But, this is unlikely to happen in practice, and therefore choosing a specific maximal R-squared will be subjective and must be guided by empirical regularity. Here, we use a rule of thumb proposed by Oster (2019, 189), which is choosing a maximal R-squared that is 1.3 times the R-squared obtained from the regression model that controls for background characteristics, birth month fixed effects, birth year fixed effects, and

residential city in 1920 fixed effects (which is 0.0278). Based on this, we use a maximal R-squared of 0.04 (i.e., $0.0278 \times 1.3 = 0.03614 \approx 0.04$). Using this maximal R-squared, Oster's method produces a bias-adjusted treatment-effect estimate of -0.2238 (shown in Panel A). In comparison, we hereafter refer to the treatment-effect estimate, which is -0.2504, as the original estimate, obtained from the regression model controlling for background characteristics, birth month fixed effects, birth year fixed effects, and residential city in 1920 fixed effects. According to Oster (2019), assuming selection on unobservables is as important as selection on observables and if the interval formed by the original estimate and the bias-adjusted estimate excludes zero, then the original estimate will be deemed as robust. In our case the interval is [-0.2504, -0.2238], which excludes zero. Therefore, we deem our original estimate as robust. This robustness can also be justified if the bias-adjusted estimate falls into the confidence interval of the original estimate. Panel A shows that the bias-adjusted estimate (-0.2238) falls into that interval, indicating that our original estimate is robust to selection on unobservables. Panel A also shows that this robustness can be justified for a wide range of maximal R-squared, going from 0.04 to 0.10.[3] In Panel B we conduct another robustness check also proposed by Oster (2019). Specifically, this method calculates the degree of selection on unobservables relative to observables, which Oster (2019) refers to as the coefficient of proportionality.[4] This calculation assumes a value for the maximal R-squared and also assumes that the true treatment effect is zero. The calculated coefficient shows how large selection on unobservables has to be in order to produce a treatment effect of zero (i.e., completely eliminating the observed effect of in utero exposure to the 1918 influenza pandemic on old-age mortality). For example, in the case of a maximal R-squared set to be 0.04, the calculated coefficient of proportionality turns out to be 8.5160, meaning that the unobservables would need to be eight to nine times as important as the observables in order to produce a treatment effect of zero. Oster (2019) recommends that equal selection (i.e., the coefficient of proportionality being equal to one) should be used as the upper bound on the coefficient of proportionality. Based on this upper bound,

[3] Note that the R-squared in the regression model that does not control for background characteristics and birth month fixed effects (i.e., only controlling for birth year fixed effects and residential city in 1920 fixed effects) is 0.0125. This R-squared increases to 0.0278 by about 122.4%, once the model controls for background characteristics and birth month fixed effects. Results in Panel A also indicates that the original estimate can be viewed as robust to the inclusion of unobservables into the regression model that raises the R-squared from 0.04 to 0.10 by about 150%.

[4] In Oster (2019), this coefficient of proportionality is referred to as "delta," which is the ratio of selection on unobservables over selection on observables. In that paper, selection on observables (or unobservables) is defined by the covariance between the treatment dummy variable and observables (or unobservables) divided by the variance of observables (or unobservables).

the observed effect will be viewed as robust if the coefficient of proportionality is greater than one. In Panel B and in the range of maximal R-squared previously discussed (regarding the results shown in Panel A), we see that all calculated coefficients are greater than one. This result again suggests that the original estimate could be viewed as robust to selection on unobservables.

DID Analysis.—Appendix Tables A6 and A7 uses a difference-in-differences (DID) design to remove the impact of any potential "one-year younger" survival disadvantage (e.g., needing to live longer to reach age 65). Specifically, we compare the 1919 versus 1918 longevity gap with the analogous gap for an earlier adjacent-cohort pair (i.e., 1915 versus 1914). In low-intensity cities (Appendix Table A6), the DID estimates for $((1919-1918)-(1915-1914))$ are statistically significant, similar to the main results reported in Panel A of Table 2, and remain stable when adding background controls, residential city in 1920 fixed effects, and birth month fixed effects. This finding reinforces our study's main conclusion that scarring effects are detectable even where pandemic intensity was modest. In contrast, in high-intensity cities (Appendix Table A7), the corresponding DID estimate is small and statistically insignificant in every specification, consistent with the interpretation that strong selection (culling) can mask (or offset) scarring in these settings. Finally, placebo DID comparisons that shift the "treated" adjacent cohorts to earlier years, that is, $((1918-1917)-(1915-1914))$ and $((1917-1916)-(1915-1914))$, show no effects in either low-intensity or high-intensity settings, indicating that the falsification test is passed.

Appendix B: Estimations Using CenSoc-Numident Data

An alternative CenSoc mortality file is the Social Security Administration (SSA)'s Numerical Identification System (Numident) file linked to the full-count 1940 U.S. Census, commonly referred to as CenSoc-Numident (Goldstein et al. 2023). The version used in our study (and available at the time of our study) is version 3.0.[5] Like CenSoc-DMF (explained in the main text of the paper), it links individuals observed in the 1940 Census to SSA death records and covers deaths occurring at ages 65 or above (Breen, Osborne and Goldstein 2023). It also includes a unique identifier that allows researchers to merge mortality outcomes back onto individual-level variables in the publicly accessible 1940 Census data provided by IPUMS USA (explained in the

[5] For details, see https://censoc.berkeley.edu/data/ (accessed in September 2025).

main text of the paper). CenSoc-Numident includes both males and females, and it contains about 7.0 million person-level records.

CenSoc-Numident was constructed using a deterministic record-linkage algorithm described by Breen and Osborne (2022). Unlike CenSoc-DMF, linkage for females is possible in CenSoc-Numident because the SSA's Numident file contains parental information that can be used in special cases, such as linking never-married females via their fathers' surnames. However, linkage remains infeasible for females whose surnames change (Breen, Osborne and Goldstein 2023). Similar to CenSoc-DMF, CenSoc-Numident's mortality-adjusted match rate (explained in the main text of the paper) is about 30% for males and about 32% for females. In terms of representativeness, it is also slightly skewed toward higher socioeconomic status (e.g., higher high-school completion than the general population), and Black people are under-represented in CenSoc-Numident (Breen and Osborne 2022).

A key limitation of CenSoc-Numident for studies of old-age mortality is its shorter death coverage window, 1988–2005, which is substantially shorter than that of CenSoc-DMF version 4.0 (1975–2020). This shorter window implies more truncation than in CenSoc-DMF, reducing the ability to observe old-age mortality for more recent cohorts. For example, CenSoc-Numident does not include individuals surviving beyond 2005 born in 1918 (at least 87 years old) or born in 1917 (at least 88 years old). This implies that the observed sample could be negatively selected on health relative to a setting without truncation. With such truncation, comparisons of a younger cohort and an older cohort may mechanically generate an apparent "treatment effect" since the younger cohort is observed dying at younger ages than the older cohort. This issue is reflected in the similar results obtained using CenSoc-Numident to estimate the longevity impact of pandemic exposure across specifications in both low- and high-intensity cities and for both males and females, reported in Appendix A8. This truncation concern is the main reason that our study uses the CenSoc-DMF data instead.

Appendix C: Conditional Average Treatment Effect (CATE)

This section provides additional details on the machine learning (ML) methods used to estimate individualized average treatment effects (IATE) and to construct sorted group average treatment effects (GATES) based on the estimated IATE for our heterogeneous treatment effect analysis. Both IATE and GATES are specific examples of CATE, which is defined as follows:

(1) $$CATE = \tau(\boldsymbol{x}) = E[y(1) - y(0)|\boldsymbol{x}],$$

and CATE measures the average treatment effect, $ATE = E[y(1) - y(0)]$, as a function of $\boldsymbol{x}$. In our study, y denotes the outcome; d denotes the treatment indicator (1/0), equal to one for individuals born in 1919 and zero for those born in 1918; and $\boldsymbol{x}$ is a vector of background characteristics obtained from the 1920 census: dummy variables for own race, own state of birth, father's state of birth (if within the U.S.) or country of birth (if outside the U.S.) of birth, mother's state of birth (if within the U.S.) or country of birth (if outside the U.S.), father's occupation code (based on the Census Bureau occupational classification system), a second-generation immigrant father, whether the father moved outside of birthplace before child was born, an illiterate father, an illiterate mother, whether the mother was in the labor force, whether the family owned the home, birth month, and residential city in 1920; father's age when child was born; mother's age when child was born; the difference between father's and mother's ages; and birth order.

IATE is CATE at the finest level, where $\boldsymbol{x}$ uses observed values of an individual. GATES is based on sorted IATE. In our study, we divide the data into four groups, using the quantiles of IATE estimates. Specifically, the first group consists of observations with IATE estimates greater than the 75th percentile of the IATE estimates; the second group includes observations with IATE estimates lie between the 50th and 75th percentiles; the third group contains observations with IATE estimates between the 25th and 50th percentile; and the last group contains observations with IATE estimates below the 25th percentile. The group's ranking is generated using a 10-fold cross-fitting technique to avoid overfitting. To examine the presence of treatment effect heterogeneity, we test whether individuals' characteristics, such as race and education, on average are different between groups 1 and 4 (i.e., the top and bottom groups).

The estimation of IATE uses an augmented inverse-probability weighting (AIPW) estimator together with this ML algorithm—generalized random forest (Athey, Tibshirani and Wager 2019)—applied to the following model:

(2) $$y(1) = g_1(\boldsymbol{x}) + \epsilon_1, E(\epsilon_1|\boldsymbol{x}) = 0,$$

(3) $$y(0) = g_0(\boldsymbol{x}) + \epsilon_0, E(\epsilon_0|\boldsymbol{x}) = 0,$$

(4) $$d = f(\boldsymbol{x}) + u, E(u|\boldsymbol{x}) = 0,$$

where ϵ_1, ϵ_0 and u are error terms. The AIPW estimators of $y(1)$ and $y(0)$ are:

(5)

$$y(1)_{AIPW} = g_1(\boldsymbol{x}) + \frac{1\{d = 1\}(y - g_1(\boldsymbol{x}))}{f(\boldsymbol{x})}$$

(6)

$$y(0)_{AIPW} = g_0(x) + \frac{1\{d = 0\}(y - g_0(\boldsymbol{x}))}{1 - f(\boldsymbol{x})}$$

Denote the estimated IATE as follows:

(7) $$\hat{\Gamma} = \widehat{y(1)}_{AIPW} - \widehat{y(0)}_{AIPW}.$$

For IATE, we solve $\tau(\boldsymbol{x})$ in

(8)

$$\sum_{i=1}^{N} \left[\alpha_i(\boldsymbol{x}) \left(\hat{\Gamma}_i - \tau(\boldsymbol{x})\right)\right] = 0$$

where $\alpha_i(\boldsymbol{x})$ is the local weight (obtained by generalized random forest) that gives more weight to observations close to $\boldsymbol{x}$, and we use $\hat{\Gamma}$ as the dependent variable in a prediction setting that implements generalized random forest (Athey, Tibshirani and Wager 2019) and double/debiased ML with 10-fold cross-fitting (Chernozhukov et al. 2018; Kennedy 2023; Knaus 2022; Semenova and Chernozhukov 2021). To obtain GATES estimates, we regress $\hat{\Gamma}$ on a discrete-valued group indicator to obtain the mean of $\hat{\Gamma}$ within each group.

REFERENCES

Athey, S., J. Tibshirani and S. Wager (2019). "Generalized Random Forests." *The Annals of Statistics* 47 (2): 1148–1178.

Beach, B., R. Brown, J. Ferrie, M. Saavedra and D. Thomas (2022). "Reevaluating the Long-Term Impact of In Utero Exposure to the 1918 Influenza Pandemic." *Journal of Political Economy* 130 (7): 1963–1990.

Breen, C. F. and M. Osborne (2022). "An Assessment of CenSoc Match Quality." Retrieved from https://censoc.berkeley.edu/wp-content/uploads/2022/06/censoc_match_quality_report_2_1.pdf in September 2025.

Breen, C. F., M. Osborne and J. R. Goldstein (2023). "CenSoc: Public Linked Administrative Mortality Records for Individual-Level Research." *Scientific Data* 10: 802.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey and J. Robins (2018). "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal* 21 (1): C1–C68.

Goldstein, J. R., M. Alexander, C. Breen, A. M. González, F. Menares, M. Osborne, M. Snyder and U. Yildirim (2023). *CenSoc Mortality File: Version 3.0*. Berkeley: University of California.

Kennedy, E. H. (2023). “Towards Optimal Doubly Robust Estimation of Heterogeneous Causal Effects.” *Electronic Journal of Statistics* 17 (2): 3008–3049.

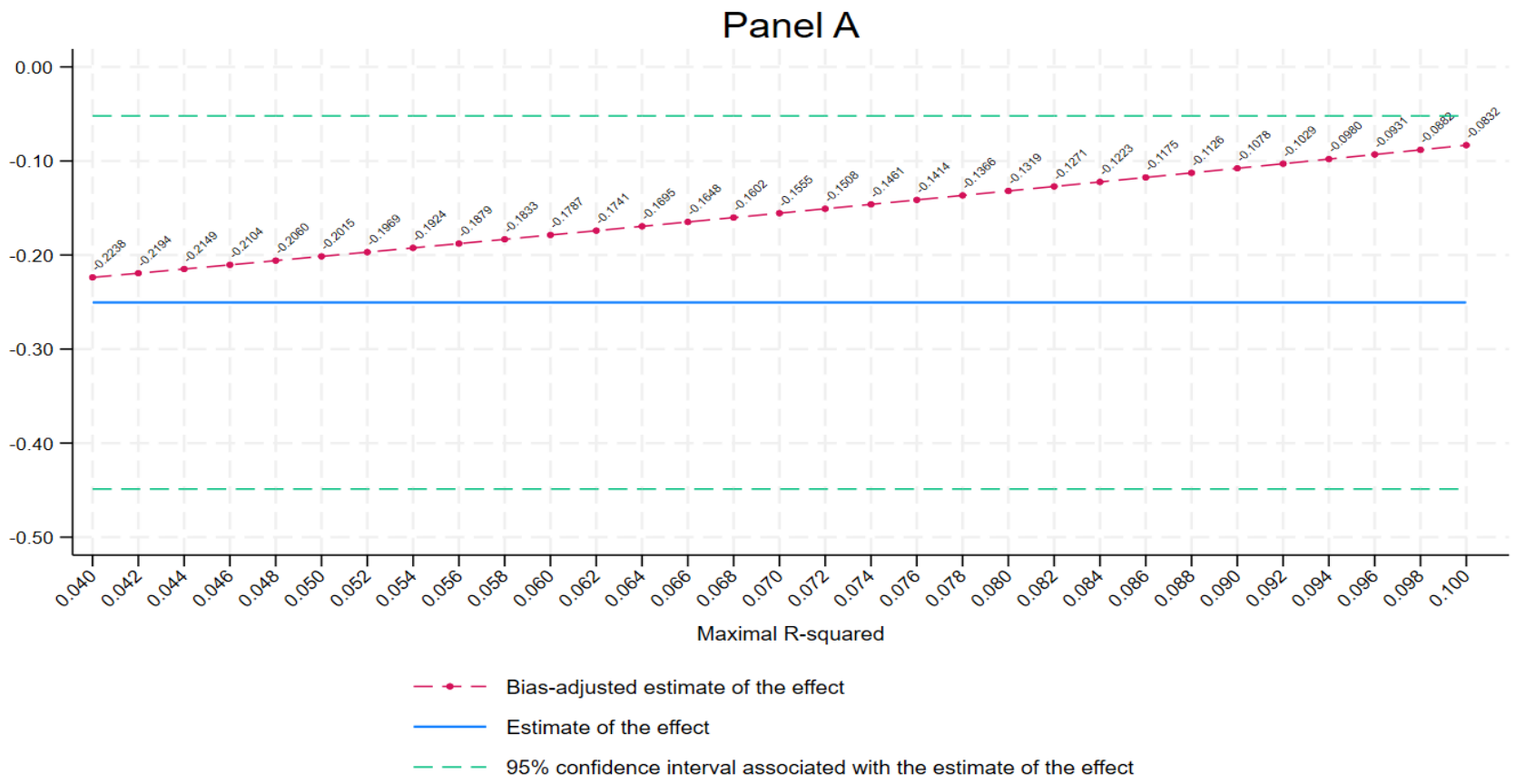
Knaus, M. C. (2022). “Double Machine Learning-Based Programme Evaluation under Unconfoundedness.” *The Econometrics Journal* 25 (3): 602–627.

Lee, D. S. (2009). “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects.” *The Review of Economic Studies* 76 (3): 1071–1102.

Lin, M.-J. and E. M. Liu (2014). “Does In Utero Exposure to Illness Matter? The 1918 Influenza Epidemic in Taiwan as a Natural Experiment.” *Journal of Health Economics* 37: 152–163.

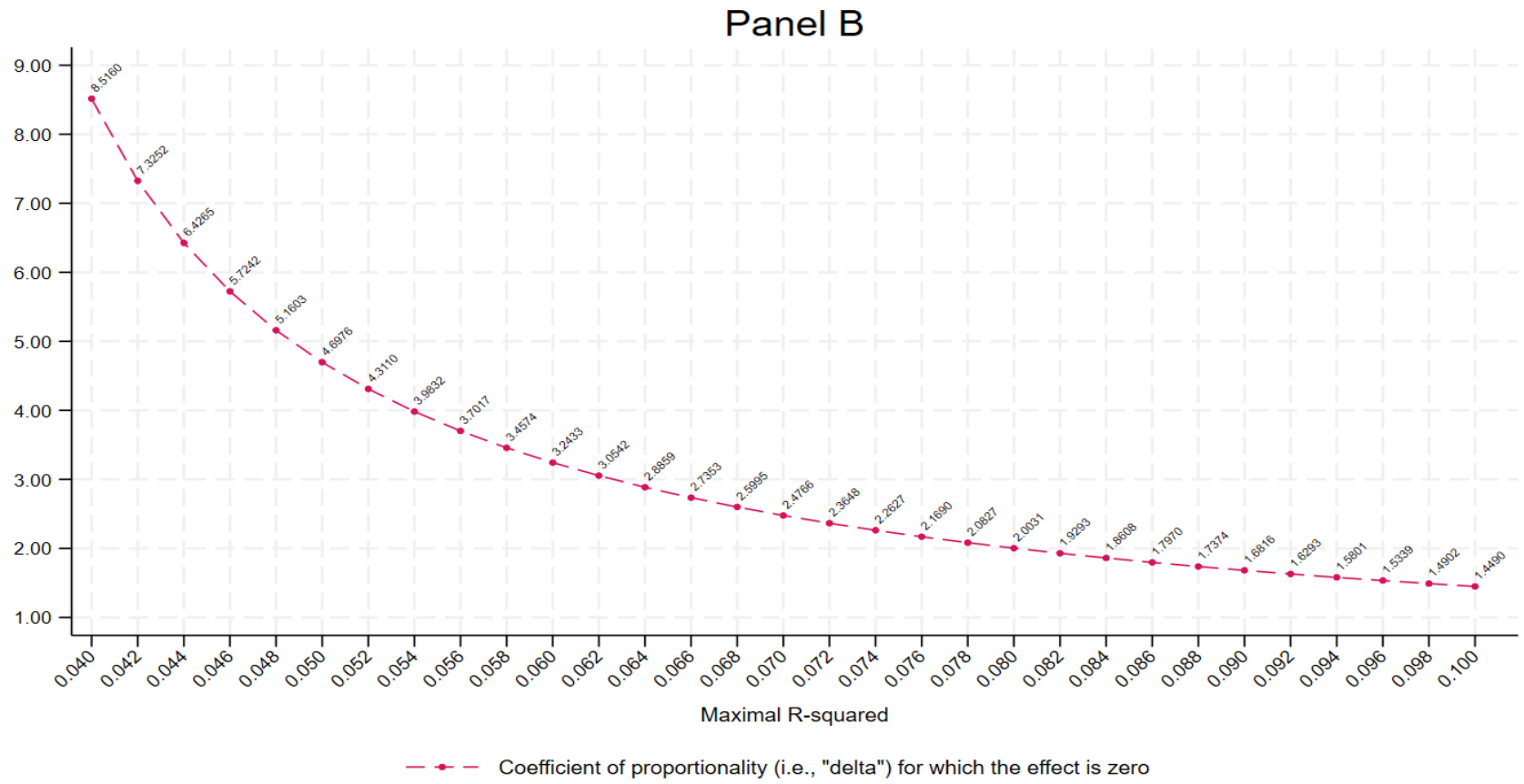
Oster, E. (2019). “Unobservable Selection and Coefficient Stability: Theory and Evidence.” *Journal of Business & Economic Statistics* 37 (2): 187–204.

Semenova, V. and V. Chernozhukov (2021). “Debiased Machine Learning of Conditional Average Treatment Effects and Other Causal Functions.” *The Econometrics Journal* 24 (2): 264–289.



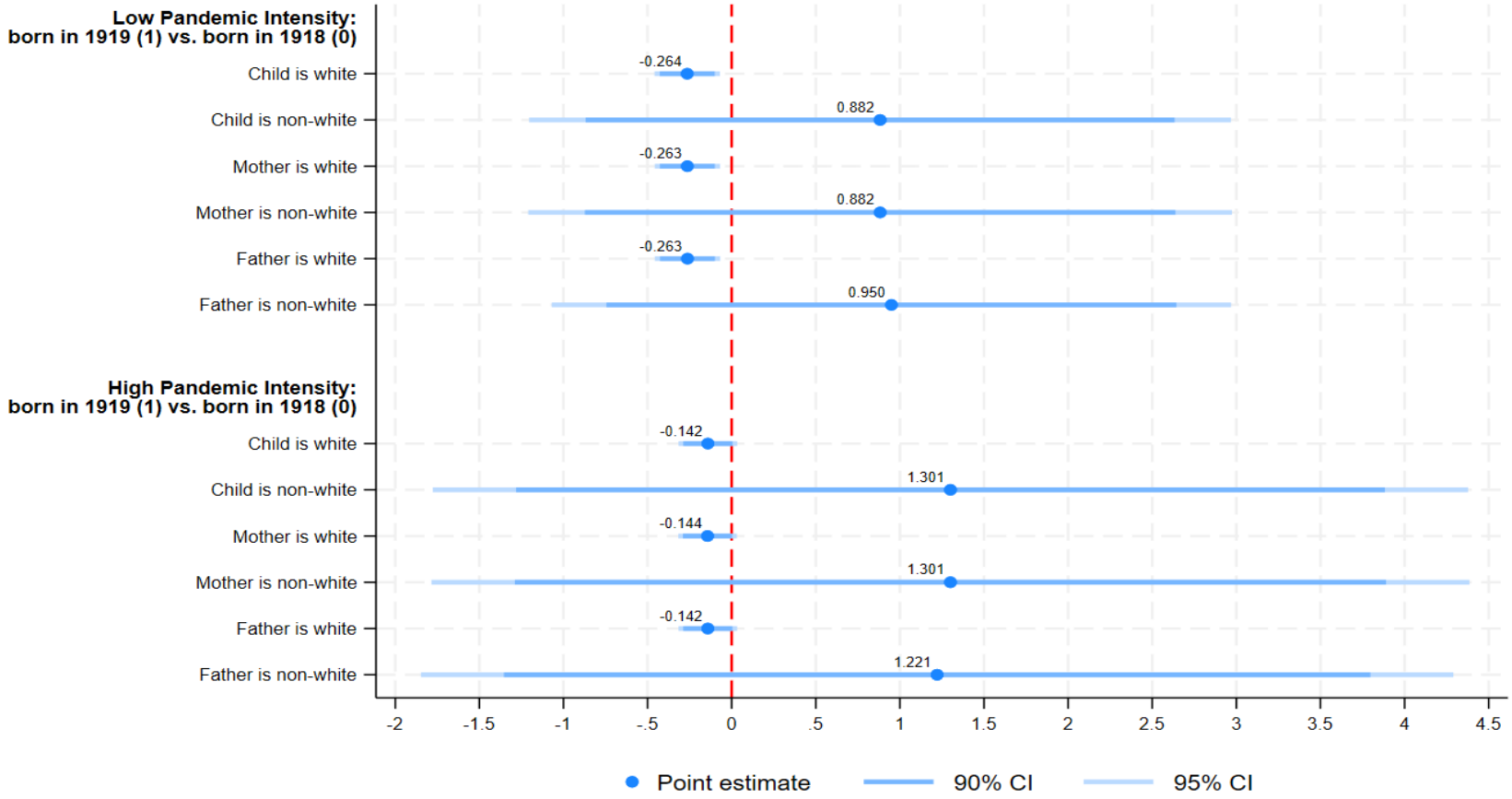
Appendix Figure A1: Robustness Checks on the Estimated Effect of In Utero Exposure to the 1918 Influenza Pandemic on Old-Age Mortality for Males in Low-Intensity Cities

Notes: The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. The median of the standardized city-level excess influenza mortality measure is used to define low-intensity cities (less than or equal to the median) and high-intensity cities (greater than the median). In Panel A, the blue line represents the estimate of the effect, which is -0.2504 and statistically significant at the 5% level. The associated confidence interval at the 95% level is also reported in Panel A. The bias-adjusted estimate of the effect is referred to as "beta star" in Oster (2019). Also in Oster (2019) the coefficient of proportionality is referred to as "delta," which is the ratio of selection on unobservables over selection on observables. In Oster (2019) selection on observables (or unobservables) is defined by the covariance between the treatment dummy variable (equal to one for those born in 1919, and zero for those born in 1918) and observables (or unobservables) divided by the variance of observables (or unobservables). Standard errors are clustered at the residential city in 1920 level.



Appendix Figure A1: Robustness Checks on the Estimated Effect of In Utero Exposure to the 1918 Influenza Pandemic on Old-Age Mortality for Males in Low-Intensity Cities (cont.)

Notes: The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. The median of the standardized city-level excess influenza mortality measure is used to define low-intensity cities (less than or equal to the median) and high-intensity cities (greater than the median). In Panel A, the blue line represents the estimate of the effect, which is -0.2504 and statistically significant at the 5% level. The associated confidence interval at the 95% level is also reported in Panel A. The bias-adjusted estimate of the effect is referred to as "beta star" in Oster (2019). Also in Oster (2019) the coefficient of proportionality is referred to as "delta," which is the ratio of selection on unobservables over selection on observables. In Oster (2019) selection on observables (or unobservables) is defined by the covariance between the treatment dummy variable (equal to one for those born in 1919, and zero for those born in 1918) and observables (or unobservables) divided by the variance of observables (or unobservables). Standard errors are clustered at the residential city in 1920 level.



Appendix Figure A2: Differences in Age at Death of Males Born in 1919 Relative to Those Born in 1918 by Race

Notes: The estimation sample uses data from CenSoc-DMF, the 1920 full-count census, the 1920–1940 linked census, the 1940–1950 linked census, and Beach et al. (2022), which provides the city-level 1918 influenza pandemic intensity measure. The median of the standardized city-level excess influenza mortality measure is used to define low-intensity cities (less than or equal to the median) and high-intensity cities (greater than the median). The figure reports point estimates and confidence intervals (CI) of the differences in age at death (measured in years) between adjacent birth cohorts. These differences are estimated by OLS using regression models that control for birth month fixed effects and residential city in 1920 fixed effects, as well as background characteristics. These background characteristics are from the 1920 census, including: dummy variables for own race, own state of birth, father's state of birth (if within the U.S.) or country of birth (if outside the U.S.) of birth, mother's state of birth (if within the U.S.) or country of birth (if outside the U.S.), father's occupation code (based on the Census Bureau occupational classification system), a second-generation immigrant father, whether the father moved outside of birthplace before child was born, an illiterate father, an illiterate mother, whether the mother was in the labor force, whether the family owned the home; father's age when child was born; mother's age when child was born; the difference between father's and mother's ages; and birth order. Standard errors are clustered at the residential city in 1920 level.

Appendix Table A1: Differences in Age at Death of Males Born in Adjacent Years in Low Pandemic Intensity Cities, Robustness Checks

	(1)	(2)	(3)	(4)
Panel A: use mean instead of median				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.244 ***	-0.227 **	-0.229 **	-0.227 **
(Standard error)	(0.086)	(0.089)	(0.090)	(0.088)
[P-value]	[0.005]	[0.011]	[0.011]	[0.011]
Panel B: cluster standard errors clustered at the county level				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.264 **	-0.246 **	-0.252 **	-0.250 **
(Standard error)	(0.103)	(0.102)	(0.102)	(0.102)
[P-value]	[0.011]	[0.017]	[0.015]	[0.015]
Panel C: use absolute median value obtained from the flu city data				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.264 ***	-0.246 **	-0.252 **	-0.250 **
(Standard error)	(0.097)	(0.102)	(0.102)	(0.101)
[P-value]	[0.007]	[0.017]	[0.014]	[0.014]
Control variables used for Panels A, B and C				
Background characteristics	No	Yes	Yes	Yes
City FE	No	No	Yes	Yes
Birth month FE	No	No	No	Yes

Notes: See Table 2 except that sample sizes in Panels A, B and C respectively are: 51,140; 51,140; 51,140. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

Appendix Table A2: Differences in Age at Death of Males Born in Adjacent Years in High Pandemic Intensity Cities, Robustness Checks

	(1)	(2)	(3)	(4)
Panel A: use mean instead of median				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.117	-0.127	-0.126	-0.126
(Standard error)	(0.114)	(0.116)	(0.115)	(0.116)
[P-value]	[0.307]	[0.276]	[0.276]	[0.280]
Panel B: cluster standard errors clustered at the county level				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.134	-0.136	-0.131	-0.131
(Standard error)	(0.084)	(0.086)	(0.084)	(0.084)
[P-value]	[0.110]	[0.113]	[0.118]	[0.122]
Panel C: use absolute median value obtained from the flu city data				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.134	-0.136	-0.131	-0.131
(Standard error)	(0.087)	(0.088)	(0.088)	(0.088)
[P-value]	[0.122]	[0.125]	[0.136]	[0.140]
Control variables used for Panels A, B and C				
Background characteristics	No	Yes	Yes	Yes
City FE	No	No	Yes	Yes
Birth month FE	No	No	No	Yes

Notes: See Appendix Table A1. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

Appendix Table A3: Effects of In Utero Exposure to the 1918 Influenza Pandemic on Age at Death for Males in Low or High Pandemic Intensity Cities, Estimated Using Lee Bounds

	(1)	(2)	(3)
Panel A: Low Pandemic Intensity			
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.250 **	-0.809 ***	-0.809 ***
(Standard error)	(0.101)	(0.093)	(0.093)
[P-value]	[0.014]	[0.000]	[0.000]
Background characteristics	Yes	Yes	Yes
City FE	Yes	Yes	Yes
Birth month FE	Yes	Yes	Yes
The group that was trimmed	None trimmed	Control group trimmed from below	Control group trimmed from below
Trimming percentage	0.0	6.0	7.4
Number of observations	51,140	49,736	49,736
Panel B: High Pandemic Intensity			
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.131	-0.803 ***	-0.803 ***
(Standard error)	(0.088)	(0.090)	(0.090)
[P-value]	[0.140]	[0.000]	[0.000]
Background characteristics	Yes	Yes	Yes
City FE	Yes	Yes	Yes
Birth month FE	Yes	Yes	Yes
The group that was trimmed	None trimmed	Control group trimmed from below	Control group trimmed from below
Trimming percentage	0.0	6.0	7.4
Number of observations	51,140	49,736	49,736

Notes: See Table 2. Column (1) uses the original estimation sample; column (2) uses the 1920 full count data to determine attrition; and column (3) uses the 1920–1940 linked census data to determine attrition. * p -value < 0.1; ** p -value < 0.05; *** p -value < 0.01.

Appendix Table A4: Differences in Age at Death of Males Born in Adjacent Years in Low-Intensity Cities, Estimated Using Alternative Definition of Pandemic Intensity

	(1)	(2)	(3)	(4)
Panel A				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.270 **	-0.246 **	-0.256 **	-0.259 **
(Standard error)	(0.114)	(0.108)	(0.110)	(0.110)
[P-value]	[0.019]	[0.024]	[0.020]	[0.019]
Panel B				
Birth cohort (1/0): 1 for 1918 and 0 for 1917	-0.059	-0.034	-0.029	-0.025
(Standard error)	(0.100)	(0.100)	(0.100)	(0.101)
[P-value]	[0.557]	[0.734]	[0.773]	[0.802]
Panel C				
Birth cohort (1/0): 1 for 1917 and 0 for 1916	-0.058	-0.061	-0.065	-0.069
(Standard error)	(0.106)	(0.108)	(0.107)	(0.107)
[P-value]	[0.583]	[0.571]	[0.545]	[0.517]
Panel D				
Birth cohort (1/0): 1 for 1916 and 0 for 1915	-0.160	-0.115	-0.106	-0.103
(Standard error)	(0.123)	(0.124)	(0.125)	(0.124)
[P-value]	[0.192]	[0.357]	[0.395]	[0.407]
Panel E				
Birth cohort (1/0): 1 for 1915 and 0 for 1914	0.006	0.024	0.029	0.029
(Standard error)	(0.126)	(0.120)	(0.120)	(0.119)
[P-value]	[0.964]	[0.838]	[0.808]	[0.807]
Control variables used for Panels A through E				
Background characteristics	No	Yes	Yes	Yes
City FE	No	No	Yes	Yes
Birth month FE	No	No	No	Yes

Notes: See Table 2 except that 1) low-intensity cities in the estimation sample are cities in the bottom 10 states based on the 1918 influenza mortality rate rank in table 1 of Garrett (2008, https://doi.org/10.20955/r.90.74-94) and high-intensity cities in the estimation sample are cities in the top 10 states based on the 1918 influenza mortality rate rank in table 1 of table 1 of Garrett (2008, https://doi.org/10.20955/r.90.74-94); 2) sample sizes in Panels A, B, C, D and E respectively are: 37,079; 39,093; 38,725; 38,233; 37,669. * p -value < 0.1; ** p -value < 0.05; *** p -value < 0.01.

Appendix Table A5: Differences in Age at Death of Males Born in Adjacent Years in High-Intensity Cities, Estimated Using Alternative Definition of Pandemic Intensity

	(1)	(2)	(3)	(4)
Panel A				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.207 *	-0.196 *	-0.178	-0.179
(Standard error)	(0.107)	(0.110)	(0.110)	(0.110)
[P-value]	[0.054]	[0.075]	[0.107]	[0.103]
Panel B				
Birth cohort (1/0): 1 for 1918 and 0 for 1917	-0.033	-0.006	0.002	0.002
(Standard error)	(0.102)	(0.106)	(0.108)	(0.108)
[P-value]	[0.748]	[0.958]	[0.989]	[0.982]
Panel C				
Birth cohort (1/0): 1 for 1917 and 0 for 1916	-0.076	-0.060	-0.076	-0.077
(Standard error)	(0.115)	(0.113)	(0.112)	(0.112)
[P-value]	[0.508]	[0.597]	[0.501]	[0.493]
Panel D				
Birth cohort (1/0): 1 for 1916 and 0 for 1915	0.098	0.121	0.110	0.108
(Standard error)	(0.117)	(0.116)	(0.118)	(0.118)
[P-value]	[0.404]	[0.299]	[0.354]	[0.364]
Panel E				
Birth cohort (1/0): 1 for 1915 and 0 for 1914	0.118	0.137	0.161	0.165
(Standard error)	(0.108)	(0.114)	(0.115)	(0.115)
[P-value]	[0.276]	[0.231]	[0.160]	[0.150]
Control variables used for Panels A through E				
Background characteristics	No	Yes	Yes	Yes
City FE	No	No	Yes	Yes
Birth month FE	No	No	No	Yes

Notes: See Appendix Table A4. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.

Appendix Table A6: Effects of In Utero Exposure to the 1918 Influenza Pandemic on Age at Death for Males in Low Pandemic Intensity Cities, Estimated by DID

	(1)	(2)	(3)	(4)
Panel A: (1919 - 1918) - (1915 - 1914)	-0.318 **	-0.298 **	-0.300 **	-0.297 **
Treated (1/0) × After (1/0)	(0.124)	(0.120)	(0.122)	(0.122)
(Standard error)	[0.011]	[0.014]	[0.015]	[0.016]
[P-value]				
Panel B: (1918 - 1917) - (1915 - 1914)				
Treated (1/0) × After (1/0)	-0.025	-0.008	-0.007	-0.009
(Standard error)	(0.091)	(0.096)	(0.096)	(0.095)
[P-value]	[0.787]	[0.937]	[0.939]	[0.923]
Panel C: (1917 - 1916) - (1915 - 1914)				
Treated (1/0) × After (1/0)	-0.167	-0.164	-0.177	-0.179
(Standard error)	(0.117)	(0.115)	(0.114)	(0.114)
[P-value]	[0.158]	[0.159]	[0.124]	[0.119]
Control variables used for Panels A, B and C				
Background characteristics	No	Yes	Yes	Yes
City FE	No	No	Yes	Yes
Birth month FE	No	No	No	Yes

Notes: See Table 2 except that: 1) the regression model uses a difference-in-differences (DID) specification described in the panel title; "Treat (1/0)" equals one for the treatment group (consisting of the two adjacent birth cohorts of later years), and zero for the control group (consisting of the birth cohorts of 1914 and 1915); "After (1/0)" equals one for the later years of the birth years constituting the treatment and control groups, and zero for the earlier years of the birth years constituting the treatment and control groups; 2) sample sizes in Panels A, B and C respectively are: 53,326; 54,524; 54,069. * p -value < 0.1; ** p -value < 0.05; *** p -value < 0.01.

Appendix Table A7: Effects of In Utero Exposure to the 1918 Influenza Pandemic on Age at Death for Males in High Pandemic Intensity Cities, Estimated by DID

	(1)	(2)	(3)	(4)
Panel A: (1919 - 1918) - (1915 - 1914)				
Treated (1/0) × After (1/0)	-0.107	-0.122	-0.119	-0.122
(Standard error)	(0.139)	(0.144)	(0.143)	(0.144)
[P-value]	[0.444]	[0.397]	[0.408]	[0.395]
Panel B: (1918 - 1917) - (1915 - 1914)				
Treated (1/0) × After (1/0)	-0.168	-0.160	-0.156	-0.153
(Standard error)	(0.149)	(0.149)	(0.149)	(0.149)
[P-value]	[0.261]	[0.283]	[0.298]	[0.304]
Panel C: (1917 - 1916) - (1915 - 1914)				
Treated (1/0) × After (1/0)	-0.056	-0.084	-0.085	-0.090
(Standard error)	(0.116)	(0.114)	(0.114)	(0.114)
[P-value]	[0.628]	[0.461]	[0.458]	[0.431]
Control variables used for Panels A, B and C				
Background characteristics	No	Yes	Yes	Yes
City FE	No	No	Yes	Yes
Birth month FE	No	No	No	Yes

Notes: See Table 2 except that: 1) the regression model uses a difference-in-differences (DID) specification described in the panel title; "Treat (1/0)" equals one for the treatment group (consisting of the two adjacent birth cohorts of later years), and zero for the control group (consisting of the birth cohorts of 1914 and 1915); "After (1/0)" equals one for the later years of the birth years constituting the treatment and control groups, and zero for the earlier years of the birth years constituting the treatment and control groups; 2) sample sizes in Panels A, B and C respectively are: 49,315; 50,673; 50,569. * p -value < 0.1; ** p -value < 0.05; *** p -value < 0.01.

Appendix Table A8: Effects of In Utero Exposure to the 1918 Influenza Pandemic on Age at Death in Low or High Pandemic Intensity Cities, Estimated Using Numident Data

	(1)	(2)	(3)	(4)
Panel A: Low pandemic intensity, full sample				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.875 ***	-0.877 ***	-0.879 ***	-0.864 ***
(Standard error)	(0.033)	(0.037)	(0.037)	(0.037)
[P-value]	[0.000]	[0.000]	[0.000]	[0.000]
Panel B: Low pandemic intensity, male sample				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.897 ***	-0.884 ***	-0.884 ***	-0.868 ***
(Standard error)	(0.042)	(0.044)	(0.044)	(0.043)
[P-value]	[0.000]	[0.000]	[0.000]	[0.000]
Panel C: Low pandemic intensity, female sample				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.864 ***	-0.866 ***	-0.867 ***	-0.854 ***
(Standard error)	(0.055)	(0.057)	(0.058)	(0.058)
[P-value]	[0.000]	[0.000]	[0.000]	[0.000]
Panel D: High pandemic intensity, full sample				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.928 ***	-0.932 ***	-0.938 ***	-0.931 ***
(Standard error)	(0.038)	(0.039)	(0.039)	(0.040)
[P-value]	[0.000]	[0.000]	[0.000]	[0.000]
Panel E: High pandemic intensity, male sample				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.945 ***	-0.930 ***	-0.932 ***	-0.923 ***
(Standard error)	(0.051)	(0.051)	(0.052)	(0.052)
[P-value]	[0.000]	[0.000]	[0.000]	[0.000]
Panel F: High pandemic intensity, female sample				
Birth cohort (1/0): 1 for 1919 and 0 for 1918	-0.928 ***	-0.930 ***	-0.936 ***	-0.930 ***
(Standard error)	(0.064)	(0.063)	(0.063)	(0.063)
[P-value]	[0.000]	[0.000]	[0.000]	[0.000]
Control variables used for Panels A through E				
Background characteristics	No	Yes	Yes	Yes
City FE	No	No	Yes	Yes
Birth month FE	No	No	No	Yes

Notes: See Table 2 except that: 1) the CenSoc-DMF data are replaced by the Numident data; 2) in the full sample, female (1/0) is controlled for; 3) sample sizes in Panels A, B, C, D, E and F respectively are: 116,227; 63,881; 52,346; 116,227; 63,881; 52,346. * p-value < 0.1; ** p-value < 0.05; *** p-value < 0.01.