# The Long Shadow of the Past: Early-Life Disease Environment and Later-Life Mortality*

Hamid Noghanibehambari†

Jason Fletcher‡

**Abstract**

A recently growing literature evaluates the influence of early-life conditions on life-cycle health and mortality. This paper extends this literature by estimating the associations between birth-state infant mortality rates experienced during early life (as a proxy for general disease environment, health-care access, and nutrition) and later-life old-age mortality rates. Using the universe of death records in the US over the years 1979-2020 and implementing two-way fixed effect models, we find positive and significant associations. Back-of-the-envelope calculations suggest that the sharp reductions in infant mortality rates across cohorts of the final sample (1915-1940) are associated with a roughly 26% drop in mortality rates during old age, conditional on survival up to age 55. Further, we find that these associations are larger for older ages and are larger in magnitudes at higher levels of infant mortality rates. Further, we provide empirical evidence to argue that deteriorations in education-income profile and increases in disability are likely mechanism channels.

**Keywords**: Mortality, Infant Mortality, Early-Life Exposures

**JEL Codes**: I18, J13, N31, N32

* The authors claim that they have no conflict of interest to report. The authors would like to acknowledge financial support from NIA grants (R01AG060109, R01AG076830).

† College of Business, Austin Peay State University, Marion St, Clarksville, TN 37040, USA
Email: noghanih@apsu.edu, Phone: +1-806-620-1812

‡ La Follette School of Public Affairs, University of Wisconsin-Madison, 1225 Observatory Drive, Madison, WI 53706-1211, USA
Email: jason.fletcher@wisc.edu

## 1. Introduction

There is a growing literature that examines the role of early-life environment in explaining the disparities in later-life outcomes. Empirical studies in various settings provide suggestive evidence that policy and environmental exposures experienced by infants and children may have a lingering influence across their life course (Almond et al., 2018; Almond & Currie, 2011a, 2011b; Barker, 1994, 1995; Currie & Rossin-Slater, 2013). For instance, studies point to the relevance of disease environment experienced during in-utero, early-life, and childhood for children's health and their later-life outcomes. Case & Paxson (2009) document that those born in census regions with higher infant mortality rates are more likely to reveal impaired cognitive functioning during old age. Bozzoli et al. (2009) employ data from several developed countries and show that individuals who were born in countries with higher infant mortality reveal lower height across the life course. Although the extent of the outcomes explored in these studies is limited, they point to declining health outcomes for individuals with higher exposure to adversities in early-life with potential spillover and effects on other outcomes.

The current study extends this literature by evaluating the correlations between area level infant mortality rates at the level of the birth-state-year, as a proxy for the general disease environment, undernutrition, and lack of healthcare access, and later-life old-age age-specific mortality rates. We find that those born in states with higher infant mortality rates reveal higher mortality rates in ages past 54. Further, we evaluate the heterogeneity of these correlations across cohorts and find that the scarring influence of early-life disease environments is more pronounced for earlier cohorts who were exposed to higher infant mortality regimes. We also find that these correlations are significant for several causes of death in old age, including infectious diseases, malignant neoplasm diseases, and cardiovascular diseases. The rest of the paper is organized as follows. Section 2 introduces data sources. Section 3 discusses the econometric method. Section 4 reviews the results. Finally, we conclude the paper in section 5.

## 2. Data Sources

The primary source of data is vital statistics death certificates extracted from NCHS (2020). The NCHS data covers the universe of deaths in the US. Starting from 1979, the NCHS reports the state-of-birth of decedents, a required identifier in our setting. The data also contains the age, race, ethnicity, and state-year of death. We use age-at-death and year-of-death to infer birth cohort.

We use the restricted version of this data which covers death records up to 2020. The initial sample contains roughly 88 million death records.

Infant mortality data is extracted from Bailey et al. (2016). The data contains infant death data at the county level from 1915 to 2007. We aggregate the data at the state-level since the NCHS sample reports birth-states. To calculate mortality rates, we collapse the NCHS data by birth-year, birth-state, current-year, current-state, and age. We then calculate mortality rates as follows:

$$MR_{bcsta} = \frac{N_{bcsta}}{POP_{sta}} \times 100{,}000$$

Where $MR$ is the mortality rate of cohort $c$ born in state $b$ died at age $a$ in state $s$ in year $t$. The denominator is the aggregate population by state, year, and age taken from SEER (2019).[4] We then merge this data with the infant mortality data of Bailey et al. (2016) based on birth-state and birth-year. We focus on deaths between ages 55-89 for those born between 1915-1940[5] and died between 1979-2020 in 24 states with complete data, so that our final sample is of approximately 24M decedents (see Appendix A for further details).

We also use decennial census data from 1910-1940 to calculate a series of birth-state covariates. We linearly interpolate these covariates for inter-decennial years. Moreover, in the section that explores mechanism channels, we use decennial census data for the years 1980-2000 combined with the American Community Survey data for the years 2001-2020, years that overlap with the NCHS data. These data are extracted from Ruggles et al. (2020).

Summary statistics of the final sample are reported in the left and right panels of Table 1 for birth cohorts of 1915-1927 and 1928-1940, respectively.[6] The average age-specific mortality

[4] In the formula for calculating mortality rate, the denominator and numerator have different levels of aggregation. The reason is that the SEER (2019) data does not contain information of population based on birth state and birth year and only reports population estimates based on current year and current state. In Appendix E, we examine the robustness of the results using the population estimates from decennial censuses combined with the American Community Survey data to replace the denominator. The disadvantage is that the data does not cover inter-decennial years and the population estimates are based on random samples rather than the full population as in SEER (2019). The results, reported in Appendix Table E-1, almost identical to the main results of the paper.

[5] Appendix D shows that including later-joining states—those that entered the federal death registration system only after 1920—yields smaller coefficients (0.02 vs. 0.03), while restricting to states with complete data for 1915–1940 produces larger effects, consistent with higher infant mortality levels and older ages at death amplifying the estimates.

[6] The process of constructing the final sample is explained in detail in Appendix A.

rate of the 1915-1927 (1928-1940) cohorts is 142.5 (98.3) per 1,000. Similarly, the average infant mortality rate is 78.4 and 55.9 per 1,000 for the 1915-1927 and 1928-1940 cohorts, respectively.

Figure 1 depicts the geographic distribution of states in the final sample based on the quartiles of infant mortality rate over the years 1915-1940 (top panel) and quartiles of age-specific death rates over the years 1979-2020 (bottom panel). The time-series evolution of mortality rates and infant mortality rates for nine Census divisions are depicted in the top and bottom panels of Figure 2, respectively.

## 3. Econometric Method

Our identification strategy compares age-specific death rates of cohorts born in states with higher infant mortality rates to those born in states with lower infant mortality rates, net of fixed effects and controls. Specifically, we implement panel data fixed effect models as follows:

$$MR_{brcst} = \alpha_0 + \alpha_1 IMR_{bc} + \alpha_2 X_{bcst} + \alpha_3 Z_{bc} + \eta_{cr} + \gamma_b + \xi_{st} + \varepsilon_{brcst} \quad (1)$$

Where $b$ indexes birth-state, $r$ census region of birth, $c$ birth-cohort, $s$ death-state, and $t$ death-year. Variable $IMR$ represents infant mortality rate for cohort $c$ born in state $b$. In $X$, we include share of females and nonwhites from the NCHS data. In $Z$, we include birth-state covariates, including share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, per capita midwives, and per capita doctors. Parameter $\gamma$ represents birth-state fixed effects. Parameter $\eta$ represents birth cohort fixed effects interacted with census region of birth fixed effects. Cohort fixed effects account for secular trends of mortality outcomes across cohorts. The interaction of cohort and region fixed effects accounts for differential convergence in the mortality of cohorts across different census regions (shown in Figure 2). To absorb all contemporaneous unobserved state-level confounders that vary over time, we include the interaction set of death-state and death-year fixed effects, represented by $\xi$. Finally, $\varepsilon$ is a disturbance term.

We cluster standard errors at the birth-state level to account for serial correlations in the error terms and also at the birth-cohort level to account for spatial correlations. We weight regressions by the mean state-level age-level population in the 1980 census. To ease the interpretation of coefficients, we standardize both mortality rate and infant mortality rate using the full sample's mean and standard deviation. While in the main results, we focus on a linear

relationship and employ continuous variables for both the dependent and independent variables, we show the robustness of the results to alternative functional forms in Appendix B.

## 4. Results

### 4.1. Main Results

The main results of the paper are reported in Table 2. We report the parsimonious results of a regression that only includes birth-state and birth-year-by-birth-region fixed effects in column 1. We then add death-state-by-death-year fixed effects into the regression in column 2. In column 3, we add birth-state level controls. The fully parameterized regression suggests that a one-standard-deviation rise in infant mortality at state-year-of-birth is associated with a 0.03 standard deviation higher age-specific mortality rate later in life (past age 54). To put this number into perspective, we use the reduction in infant mortality rate between the years 1920-2000. During this period, infant mortality dropped from 84.1 to 5.3 per 1,000, a decrease of about 4.4 times its standard deviation in the final sample. Therefore, the reduction in later-life age-specific mortality rate associated with this reduction is about 14.6 percent of a standard deviation change, roughly 75.6 lower age-specific deaths per 1,000.[7]

These results are in line with several studies that suggest the disease burden in early-life is associated with adverse life-cycle health outcomes (Bozzoli et al., 2009; Case & Paxson, 2009). These findings are possibly in contrast with the results of Haywarda et al. (2016) who use Finnish data and show that child mortality rate during the early years of life has no association with mortality rates between ages 15-50, however, we focus on deaths occurring past age 54.

To investigate such heterogeneities by age, we replicate the main results for each single age-at-death group and report the estimated coefficients in Figure 3. The coefficients are relatively small and insignificant up to age 59. They rise in magnitude from age 60 and become statistically significant. There is an almost constant rise in the magnitude of coefficients from age 60 to age 72. They stabilize in magnitude from around age 73 up to age 80. The set of coefficients for the age range 81-89 reveals the largest impacts, although in several cases they are statistically insignificant.

[7] Appendix C provides heterogeneity analyses results based on age, cohort, gender, and race.

The effects could also reveal heterogeneity by timing of exposure across childhood ages. To examine this possibility, we estimate a lead–lag specification relating later-life mortality to birth-state infant mortality measured in the years surrounding birth. The results are reported and discussed in Appendix F. The estimates show no evidence of differential pre-trends, as coefficients on infant mortality in the years prior to birth are small and statistically insignificant. In contrast, exposure in the year of birth is positively and significantly associated with higher later-life mortality, with a smaller and less persistent effect in the year immediately following birth. Estimates for exposures at later postnatal ages are small in magnitude and statistically indistinguishable from zero. Together, these patterns indicate that in utero and very early postnatal exposure to adverse disease environments are the most sensitive periods shaping long-run mortality risk.

### 4.2. Heterogeneity by Cause-of-Death

Disease exposure during early-life and specifically during prenatal development could result in later-life mortality risks through several channels, including changes in fetal programming. For instance, disease exposure may lead to malnutrition and alterations in the metabolism of glucose and insulin in the body with a permanent and long-lasting stamp. This adaptation could be advantageous in an environment with limited nutrition during later stages of life but can elevate the risks of cardiovascular disease and diabetes once the nutritional burden is removed (Barker, 1997; Myrskylä et al., 2013). Furthermore, early-life disease exposure may lead to infections that suppress the immune system, making it more susceptible to later-life infections. Alternatively, it may stimulate the immune system in a way that affects chronic conditions. Ochoa et al. (2004). To examine these channels, we examine the correlations across causes of death. The results are reported in Table 3.

We find mixed evidence across different causes of death. For instance, we find significant associations for mortality due to infectious diseases (column 1), malignant neoplasm diseases (column 2), cardiovascular diseases (column 4), and influenza, pneumonia, peptic ulcer, nephritis (column 5). On the other hand, we do not find significant coefficients for death due to diabetes and the category of other diseases (columns 3 and 6, respectively).

Early-life conditions, particularly the disease environment, have been shown to influence susceptibility to certain diseases in later life, including infectious and cardiovascular diseases.

(Boekelheide et al., 2012; Moore et al., 2006; Van Den Berg et al., 2011). This literature supports our findings in columns 1-5.

Further, we also observe significant associations with mortality due to external causes. One explanation is that exposure to a higher disease burden during infancy changes the life-course trajectory in a way that increases the prevalence of disability (documented in section 4.3) which is shown to influence accident-injury-related mortality (Landes et al., 2021; Patja et al., 2001). Moreover, given the growing evidence on the role of early-life conditions on later-life mental health outcomes (Zheng et al., 2021), it is not surprising to observe a significant association between early-life disease environment and suicide mortality later in life (column 8). Finally, we document the associations across different causes of death for each single age-at-death in Appendix Figure C-3 and Appendix Figure C-4.

### 4.3. Mechanism Channel

Several studies suggest that improvements in education, income, and socioeconomic status lead to increases in longevity and are associated with mortality gains (Chetty et al., 2016; Fletcher, 2015; Lleras-Muney, 2005; Meghir et al., 2018). On the other end, several studies document a link between early-life factors and later-life education-income profile (Almond, 2006; Fletcher, 2018c; Parman, 2015). Therefore, one might expect that exposure to disease environments during early-life affects later-life mortality through changes in education, income, and measures of socioeconomic status. Since the NCHS data does not report these outcomes, we use alternative data sources to explore these potential mechanism channels. In so doing, we use decennial censuses between the years 1980-2000 combined with the American Community Survey for the years 2001-2020. The coverage of these selected years matches with the death window coverage of the NCHS data.

We implement similar sample selection and regressions as equation 1. The results are reported in Table 4. We observe significant reductions in socioeconomic score (column 1), increases in the probability of low education (columns 2-3), reductions in total personal income (column 4), and increases in disability outcomes (columns 5-8). For instance, a one-standard-deviation rise in birth-state infant mortality rate is associated with a 1.8 percent drop in personal income. It also increases the probability of having less than high school and less than 12 years of

education by 1.1 and 1.8 percentage points, equivalent to roughly 14.5 and 16.8 percent change from the mean of the outcomes, respectively.

Consistent with the literature linking old-age disability outcomes to early-life experiences and the research documenting disability-related mortality outcomes (Freedman et al., 2008; Laditka & Laditka, 2019; Pongiglione et al., 2016), we observe significant increases in various measures of disability due to exposure to a harsher disease environment in early life. For instance, a one-standard-deviation increase in birth-state infant mortality rate is associated with a 21, 38, and 39 basis-points increase in the likelihood of cognitive disability, ambulatory difficulty, and independent living difficulty, respectively. These can be translated into a change of 7.8, 6.3, and 5.9 percent with respect to the mean of the outcomes, respectively. Although the coefficient of vision-or-hearing difficulty is positive, it is statistically insignificant, which limits further interpretations (column 8).

To understand what portion of correlations can be explained by these channels, we use estimates from previous research. Lindahl (2005) investigates the effects of income on health and mortality. He estimates that an increase of 10 percent in income is associated with a reduction of about 3 percentage-points in 5-year mortality, off a mean of 0.07. This can be translated into a reduction of about 43 percent with respect to the mean. Based on column 4 of Table 4, a one-standard-deviation rise in birth-state infant mortality is correlated with a roughly 2 percent decrease in total personal income. Based on Lindahl (2005)'s estimates, this drop in income suggests an increase in mortality by about 8.6 percent. On the other hand, a one-standard-deviation rise in birth-state infant mortality rate is associated with a 0.033 standard-deviation change in mortality (column 3 of Table 2). The mean and standard deviation of the mortality rate in the final sample are 1.19 and 5.18, respectively (per 1,000). Therefore, a 0.033 standard-deviation change in mortality is roughly a change of 14.3 percent with respect to the mean. The implied change based on Table 4 and Lindahl (2005) is roughly 60 percent of the reduced-form estimates of the main results ($\frac{8.6}{14.3}$). Therefore, a considerable portion of the effects can be explained by reductions in income (and arguably education).

## 5. Conclusion

Understanding the determinants of health and mortality outcomes is important for policymakers as they provide solutions to improve these outcomes. A recently developed and

growing literature points to the relevance of early-life and childhood environments and conditions for life-cycle health outcomes (Almond et al., 2018; Almond & Currie, 2011a, 2011b). The current study joined this literature and investigated the associations between state-level infant mortality experienced during in-utero and early-life, as a measure of general disease environment, undernutrition, and lack of healthcare access, on later-life old-age age-specific mortality rates. We employed the universe of death records in the US over the years 1979-2020 and implemented two-way fixed effect models to examine these associations.

Conditional on a full set of fixed effects and controls, a one-standard-deviation higher infant mortality rate (roughly equivalent to the difference in infant mortality rate in 1930 and 1940) is associated with an increase of 0.03 standard deviation in age-specific mortality rate, conditional on survival up to age 55 (roughly 13% change with respect to the mean of mortality rate over the sample period and across all ages). However, we observed that these effects are more concentrated in older ages, specifically ages past 70, suggesting delayed influences of early-life exposures. Moreover, the heterogeneity analysis by birth cohort revealed that the scarring influences of early life disease environment are more pronounced for earlier cohorts who were exposed to higher infant mortality regimes than later cohorts. Finally, we provided evidence of negative correlations between early-life infant mortality rate and later-life education, income, and socioeconomic scores. We also found positive correlations with various measures of disability as potential contributors to mortality. We argued that a large portion of the observed associations for mortality can be explained by changes in these mediatory channels.

One way to put the results into perspective is to calculate potential changes in life expectancy associated with changes in infant mortality rates experienced during early life. Between 1915 and 1940, infant mortality rates dropped by about 58% in our final sample. With respect to the mean of the infant mortality rate, this number is equivalent to roughly a two standard deviation change. Based on the main results of Table 2, this change is associated with roughly 0.06 standard deviation change in mortality rate. The mean and standard deviation of mortality in the final sample are 1.19 and 5.18, respectively. Using these statistics, we can deduce that a two-standard-deviation change in infant mortality results in a roughly 26% drop in later-life mortality. Using simplifying assumptions, this can be converted into a change in life expectancy at age 60 of about 6 years. At the beginning of the NCHS data (year 1979), life expectancy at age 60 was 23.3 and 26.1 for males and females, respectively (Social Security, 2024). Therefore, the associated

change in life expectancy is roughly 23-25 percent of the baseline life expectancy (conditional on survival up to age 60) for these cohorts.

**References**


Aizer, A., Eli, S., Ferrie, J., & Muney, A. L. (2016). The Long-Run Impact of Cash Transfers to Poor Families. *American Economic Review*, *106*(4), 935–971. https://doi.org/10.1257/AER.20140529

Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 US population. *Journal of Political Economy*, *114*(4), 672–712. https://doi.org/10.1086/507154

Almond, D., & Currie, J. (2011a). Human capital development before age five. In *Handbook of Labor Economics* (Vol. 4, Issue PART B). Elsevier. https://doi.org/10.1016/S0169-7218(11)02413-0

Almond, D., & Currie, J. (2011b). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives*, *25*(3), 153–172. https://doi.org/10.1257/JEP.25.3.153

Almond, D., Currie, J., & Duque, V. (2018). Childhood circumstances and adult outcomes: Act II. *Journal of Economic Literature*, *56*(4), 1360–1446.

Almond, D., Currie, J., & Herrmann, M. (2012). From infant to mother: Early disease environment and future maternal health. *Labour Economics*, *19*(4), 475–483. https://doi.org/10.1016/J.LABECO.2012.05.015

Bailey, M., Clay, K., Fishback, P., Haines, M., Kantor, S., Severnini, E., & Wentz, A. (2016). U.S. County-Level Natality and Mortality Data, 1915-2007. *Inter-University Consortium for Political and Social Research*. https://doi.org/https://doi.org/10.3886/E100229V4

Barker, D. J. P. (1994). *Mothers, babies, and disease in later life*. BMJ publishing group London.

Barker, D. J. P. (1995). Fetal origins of coronary heart disease. *BMJ*, *311*(6998), 171–174. https://doi.org/10.1136/BMJ.311.6998.171

Barker, D. J. P. (1997). Maternal nutrition, fetal nutrition, and disease in later life. *Nutrition*, *13*(9), 807–813. https://doi.org/10.1016/S0899-9007(97)00193-7

Beach, B., Brown, R., Ferrie, J., Saavedra, M., & Thomas, D. (2022). Reevaluating the Long-Term Impact of In Utero Exposure to the 1918 Influenza Pandemic. *Journal of Political Economy*, *130*(7), 1963–1990. https://doi.org/10.1086/719757/SUPPL_FILE/2016445DATA.ZIP

Beach, B., Clay, K., & Saavedra, M. (2022). The 1918 Influenza Pandemic and Its Lessons for COVID-19. *Journal of Economic Literature*, *60*(1), 41–84. https://doi.org/10.1257/JEL.20201641

Boekelheide, K., Blumberg, B., Chapin, R. E., Cote, I., Graziano, J. H., Janesick, A., Lane, R., Lillycrop, K., Myatt, L., States, J. C., Thayer, K. A., Waalkes, M. P., & Rogers, J. M. (2012). Predicting later-life outcomes of early-life exposures. *Environmental Health Perspectives*, *120*(10), 1353–1361. https://doi.org/10.1289/EHP.1204934/ASSET/0CFD31DE-D438-4C6E-B684-324B8A5E629B/ASSETS/GRAPHIC/EHP.1204934.G003.JPG

Bozzoli, C., Deaton, A., & Quintana-Domeque, C. (2009). Adult height and childhood disease. *Demography*, *46*(4), 647–669. https://doi.org/10.1353/DEM.0.0079

Case, A., & Paxson, C. (2009). Early Life Health and Cognitive Function in Old Age. *American Economic Review*, *99*(2), 104–109. https://doi.org/10.1257/AER.99.2.104

Case, A., & Paxson, C. (2010). Causes and consequences of early-life health. *Demography 2010 47:1*, *47*(1), S65–S85. https://doi.org/10.1353/DEM.2010.0007

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (2016). The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*, *315*(16), 1750–1766. https://doi.org/10.1001/JAMA.2016.4226

Cormack, L., Lazuka, V., & Quaranta, L. (2024). Early-Life Disease Exposure and Its Heterogeneous Effects on Mortality Throughout Life: Sweden, 1905–2016. *Demography*, *61*(4), 1187–1210. https://doi.org/10.1215/00703370-11466677

Currie, J., & Rossin-Slater, M. (2013). Weathering the storm: Hurricanes and birth outcomes. *Journal of Health Economics*, *32*(3), 487–503. https://doi.org/10.1016/j.jhealeco.2013.01.004

Cutler, D., Deaton, A., & Lleras-Muney, A. (2006). The Determinants of Mortality. *Journal of Economic Perspectives*, *20*(3), 97–120. https://doi.org/10.1257/JEP.20.3.97

Fletcher, J. M. (2015). New evidence of the effects of education on health in the US: Compulsory schooling laws revisited. *Social Science & Medicine*, *127*, 101–107. https://doi.org/10.1016/J.SOCSCIMED.2014.09.052

Fletcher, J. M. (2018a). 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. https://doi.org/10.1080/19485565.2018.1501267

Fletcher, J. M. (2018b). Examining the long-term mortality effects of early health shocks. *Applied Economics Letters*, *26*(11), 902–908. https://doi.org/10.1080/13504851.2018.1520960

Fletcher, J. M. (2018c). The effects of in utero exposure to the 1918 influenza pandemic on family formation. *Economics & Human Biology*, *30*, 59–68. https://doi.org/10.1016/J.EHB.2018.06.004

Fletcher, J., & Noghanibehambari, H. (2024). The siren song of cicadas: Early-life pesticide exposure and later-life male mortality. *Journal of Environmental Economics and Management*, *123*, 102903. https://doi.org/10.1016/J.JEEM.2023.102903

Freedman, V. A., Martin, L. G., Schoeni, R. F., & Cornman, J. C. (2008). Declines in late-life disability: The role of early- and mid-life factors. *Social Science & Medicine*, *66*(7), 1588–1602. https://doi.org/10.1016/J.SOCSCIMED.2007.11.037

Hayward, M. D., & Gorman, B. K. (2004). The long arm of childhood: The influence of early-life social conditions on men's mortality. *Demography 2004 41:1*, *41*(1), 87–107. https://doi.org/10.1353/DEM.2004.0005

Haywarda, A. D., Rigby, F. L., & Lummaa, V. (2016). Early-life disease exposure and associations with adult survival, cause of death, and reproductive success in preindustrial humans. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(32), 8951–8956. https://doi.org/10.1073/PNAS.1519820113/SUPPL_FILE/PNAS.1519820113.SAPP.PDF

Laditka, S. B., & Laditka, J. N. (2019). An Enduring Health Risk of Childhood Adversity: Earlier, More Severe, and Longer Lasting Work Disability in Adult Life. *The Journals of Gerontology: Series B*, *74*(1), 136–147. https://doi.org/10.1093/GERONB/GBY018

Landes, S. D., Stevens, J. D., & Turk, M. A. (2021). Cause of death in adults with intellectual disability in the United States. *Journal of Intellectual Disability Research*, *65*(1), 47–59. https://doi.org/10.1111/JIR.12790

Lindahl, M. (2005). Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income. *Journal of Human Resources*, *XL*(1), 144–

168. https://doi.org/10.3368/JHR.XL.1.144

Lindeboom, M., Portrait, F., & Van Den Berg, G. J. (2010). Long-run effects on longevity of a nutritional shock early in life: The Dutch Potato famine of 1846–1847. *Journal of Health Economics*, *29*(5), 617–629. https://doi.org/10.1016/J.JHEALECO.2010.06.001

Lleras-Muney, A. (2005). The Relationship Between Education and Adult Mortality in the United States. *The Review of Economic Studies*, *72*(1), 189–221. https://doi.org/10.1111/0034-6527.00329

Mazumder, B., Almond, D., Park, K., Crimmins, E. M., & Finch, C. E. (2010). Lingering prenatal effects of the 1918 influenza pandemic on cardiovascular disease. *Journal of Developmental Origins of Health and Disease*, *1*(1), 26–34. https://doi.org/10.1017/S2040174409990031

Meghir, C., Palme, M., & Simeonova, E. (2018). Education and Mortality: Evidence from a Social Experiment. *American Economic Journal: Applied Economics*, *10*(2), 234–256. https://doi.org/10.1257/APP.20150365

Moore, S. E., Collinson, A. C., N'Gom, P. T., Aspinall, R., & Prentice, A. M. (2006). Early immunological development and mortality from infectious disease in later life. *Proceedings of the Nutrition Society*, *65*(3), 311–318. https://doi.org/10.1079/PNS2006503

Myrskylä, M., Mehta, N. K., & Chang, V. W. (2013). Early life exposure to the 1918 influenza pandemic and old-age mortality by cause of death. *American Journal of Public Health*, *103*(7). https://doi.org/10.2105/AJPH.2012.301060

NCHS. (2020). Multiple Cause-of-Death Data Files. In *National Center for Health Statistics*.

Noghanibehambari, H., & Engelman, M. (2022). Social insurance programs and later-life mortality: Evidence from new deal relief spending. *Journal of Health Economics*, *86*. https://doi.org/10.1016/J.JHEALECO.2022.102690

Noghanibehambari, H., & Fletcher, J. (2023). Long-Term Health Benefits of Occupational Licensing: Evidence from Midwifery Laws. *Journal of Health Economics*, *92*, 102807. https://doi.org/10.1016/J.JHEALECO.2023.102807

Noghanibehambari, H., & Fletcher, J. (2024). Dust to Feed, Dust to Gray: The Effect of in Utero Exposure to the Dust Bowl on Old-Age Longevity. *Demography*. https://doi.org/10.1215/00703370-11140760

Ochoa, T. J., Salazar-Lindo, E., & Cleary, T. G. (2004). Management of children with infection-associated persistent diarrhea. *Seminars in Pediatric Infectious Diseases*, *15*(4), 229–236. https://doi.org/10.1053/J.SPID.2004.07.003

Parman, J. (2015). Childhood health and sibling outcomes: Nurture Reinforcing nature during the 1918 influenza pandemic. *Explorations in Economic History*, *58*, 22–43. https://doi.org/10.1016/J.EEH.2015.07.002

Patja, K., Mölsä, P., & Iivanainen, M. (2001). Cause-specific mortality of people with intellectual disability in a population-based, 35-year follow-up study. *Journal of Intellectual Disability Research*, *45*(1), 30–40. https://doi.org/10.1111/J.1365-2788.2001.00290.X

Pongiglione, B., De Stavola, B. L., Kuper, H., & Ploubidis, G. B. (2016). Disability and all-cause mortality in the older population: evidence from the English Longitudinal Study of Ageing. *European Journal of Epidemiology*, *31*(8), 735–746. https://doi.org/10.1007/S10654-016-0160-8/TABLES/4

Ruggles, S., Flood, S., Goeken, R., Grover, J., & Meyer, E. (2020). IPUMS USA: Version 10.0 [dataset]. *Minneapolis, MN: IPUMS*. https://doi.org/10.18128/D010.V10.0

Saavedra, M. (2017). Early-life disease exposure and occupational status: The impact of yellow fever during the 19th century. *Explorations in Economic History*, *64*, 62–81. https://doi.org/10.1016/J.EEH.2017.01.003

Schmitz, L. L., & Duque, V. (2022). In utero exposure to the Great Depression is reflected in late-life epigenetic aging signatures. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(46), e2208530119. https://doi.org/10.1073/PNAS.2208530119/SUPPL_FILE/PNAS.2208530119.SAPP01.PDF

SEER. (2019). Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1975-2016). *National Cancer Institute, DCCPS, Surveillance Research Program*.

Social Security. (2024). *Life Tables for the United States Social Security Area 1900-2100*. https://www.ssa.gov/oact/NOTES/as120/LifeTables_Tbl_11.html

Van Den Berg, G. J., Doblhammer-Reiter, G., Christensen, K., den Berg, G. J., Doblhammer-Reiter, G., Christensen, K., van den Berg, G. J., Doblhammer-Reiter, G., Christensen, K., den

Berg, G. J., Doblhammer-Reiter, G., & Christensen, K. (2011). Being born under adverse economic conditions leads to a higher cardiovascular mortality rate later in life: Evidence based on individuals born at different stages of the business cycle. *Demography*, *48*(2), 507–530. https://doi.org/10.1007/s13524-011-0021-8

Van Den Berg, G. J., Lindeboom, M., Portrait, F., Berg, G. J. Van Den, Lindeboom, M., Portrait, F., den Berg, G. J., Lindeboom, M., & Portrait, F. (2006). Economic Conditions Early in Life and Individual Mortality. *American Economic Review*, *96*(1), 290–302. https://doi.org/10.1257/000282806776157740

Vu, H., Noghanibehambari, H., Fletcher, J., & Green, T. (2023). Prenatal Exposure to Racial Violence and Later Life Mortality among Males: Evidence from Lynching. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4613372

Zheng, X., Shangguan, S., Fang, Z., & Fang, X. (2021). Early-life exposure to parental mental distress and adulthood depression among middle-aged and elderly Chinese. *Economics & Human Biology*, *41*, 100994. https://doi.org/10.1016/J.EHB.2021.100994

## Tables

**Table 1 - Summary Statistics**

| | Birth Years 1915-1927 | | Birth Years 1927-1940 | |
|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. |
| Mortality Rate (per 100K) | 1.42 | 6.01 | 0.98 | 4.27 |
| Log Mortality Rate | 3.38 | 1.51 | 3.1 | 1.43 |
| Infant Mortality (per 100K) | 78.39 | 15.64 | 55.97 | 12.02 |
| Log Infant Mortality Rate | 4.36 | .19 | 4.02 | .21 |
| Number of Deaths (Numerator, 100K) | 35.28 | 179.31 | 27.64 | 137.46 |
| Population (Denominator, 100K) | .36 | .4 | .41 | .43 |
| Age | 74.33 | 9.26 | 71.61 | 9.24 |
| Female | .46 | .31 | .43 | .32 |
| Nonwhite | .06 | .19 | .08 | .21 |
| ***Birth-State Covariates:*** | | | | |
| Share of White-Collar Workers | .04 | .01 | .05 | .01 |
| Share of Farmers | .17 | .12 | .14 | .1 |
| Share of Other Occupations | .79 | .11 | .81 | .1 |
| Socioeconomic Index | 25.98 | 3.68 | 27.6 | 3.65 |
| Share of Literate | .76 | .05 | .48 | .26 |
| Share of Midwives (per 100K) | 2.84 | 2.46 | 1.15 | 1.69 |
| Share of Doctors (per 1K) | 1.49 | .38 | 1.54 | .43 |
| Observations | 308,054 | | 336,327 | |

**Table 2 - Birth State Infant Mortality and Later-Life Mortality**

| | *Outcome: Age-Specific Mortality Rate (STD)* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Birth-State Infant Mortality Rate *(STD)* | .03214*** | .03269*** | .03274*** |
| | (.00929) | (.00945) | (.0089) |
| Observations | 644,381 | 644,381 | 644,381 |
| R-squared | .02834 | .06465 | .06528 |
| Mean DV | 0.000 | 0.000 | 0.000 |
| Birth-State FE | ✓ | ✓ | ✓ |
| Region-of-Birth by Birth-Year FE | ✓ | ✓ | ✓ |
| Death-State by Death-Year FE | | ✓ | ✓ |
| Birth-State Controls | | | ✓ |

Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.

*** p<0.01, ** p<0.05, * p<0.1

**Table 3 - Heterogeneity in the Correlations by Causes of Death**

| | *Outcome: Mortality Rate (STD), Cause:* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | TB, Syphilis Other Infectious Diseases | Neoplasm Diseases | Diabetes | Cardiovascular Diseases | Influenza, Pneumonia, Peptic Ulcer, Nephritis | All other Diseases | External Causes | Suicide |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Birth-State Infant Mortality Rate (STD) | .03308** | .01899* | .00115 | .03498** | .01761* | -.00059 | .03105*** | .0299** |
| | (.01584) | (.01049) | (.00759) | (.01357) | (.00923) | (.00683) | (.00812) | (.01195) |
| Observations | 35160 | 376644 | 169153 | 385735 | 267997 | 284380 | 123935 | 51319 |
| R-squared | .37758 | .14567 | .16154 | .12841 | .14593 | .1294 | .21753 | .5134 |
| Mean DV | -0.001 | -0.000 | -0.000 | -0.000 | 0.000 | -0.000 | -0.000 | -0.000 |

Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.
*** p<0.01, ** p<0.05, * p<0.1

**Table 4 - Mechanism Channel Using Census Data 1980-2000 and American Community Survey Data 2001-2020**

| | ***Outcomes:*** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Socioeconomic Score | Education < High School | Education < 12 | Log Total Personal Income | Cognitive Disability | Ambulatory Difficulty | Independent Living Difficulty | Vision or Hearing Difficulty |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Birth-State Infant Mortality Rate (STD) | -.4299633*** | .0111587*** | .0182048*** | -.0182211*** | .0021668** | .0038622*** | .0039649*** | .0009527 |
| | (.1018682) | (.002132) | (.0024008) | (.0027712) | (.0008671) | (.000978) | (.0009832) | (.0008822) |
| Observations | 11951216 | 21555007 | 21555007 | 20212792 | 17630188 | 17630188 | 20530879 | 17630188 |
| R-squared | .0529232 | .0756103 | .1063869 | .1297212 | .026888 | .0604788 | .0661691 | .0619006 |
| Mean DV | 46.498 | 0.062 | 0.148 | 10.267 | 0.078 | 0.199 | 0.109 | 0.118 |

Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. The independent variable is standardized (STD) with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census. The income values are in 2020 dollars.
** p<0.01, ** p<0.05, * p<0.1

## Figures

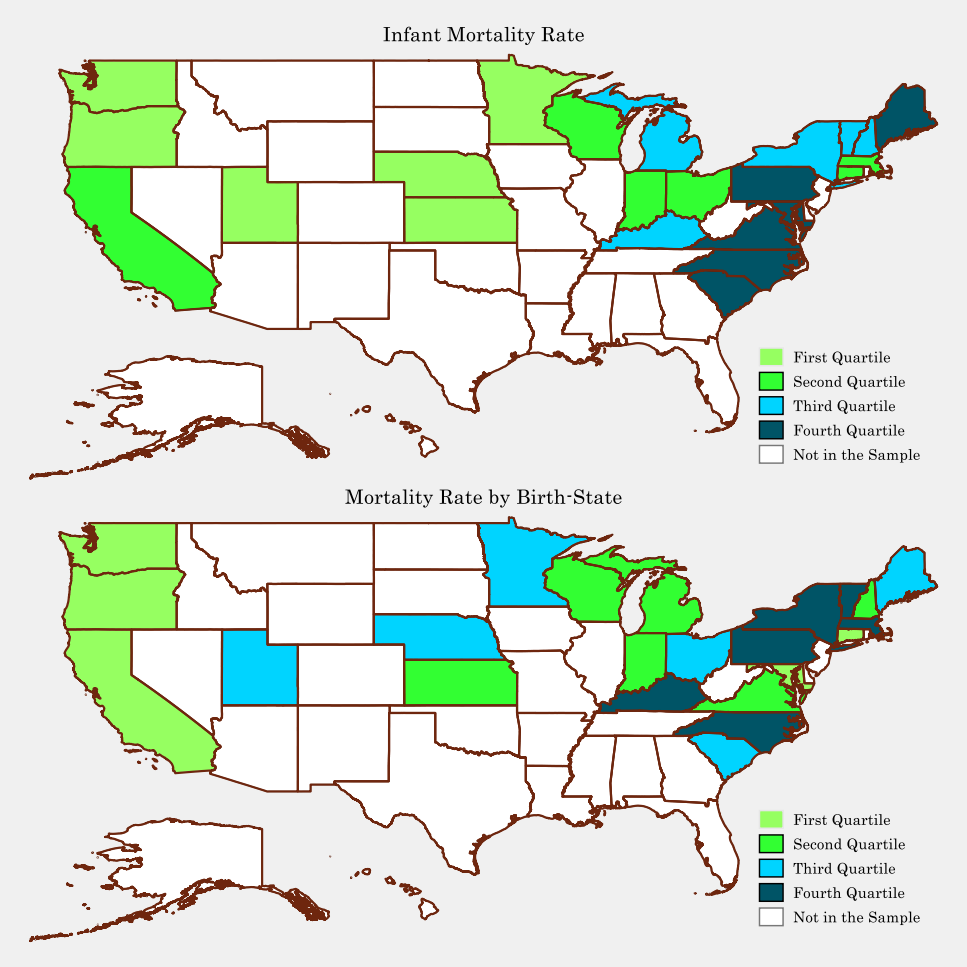



**Figure 1 - Geographic Distribution of Infant Mortality Rate by Birth States**

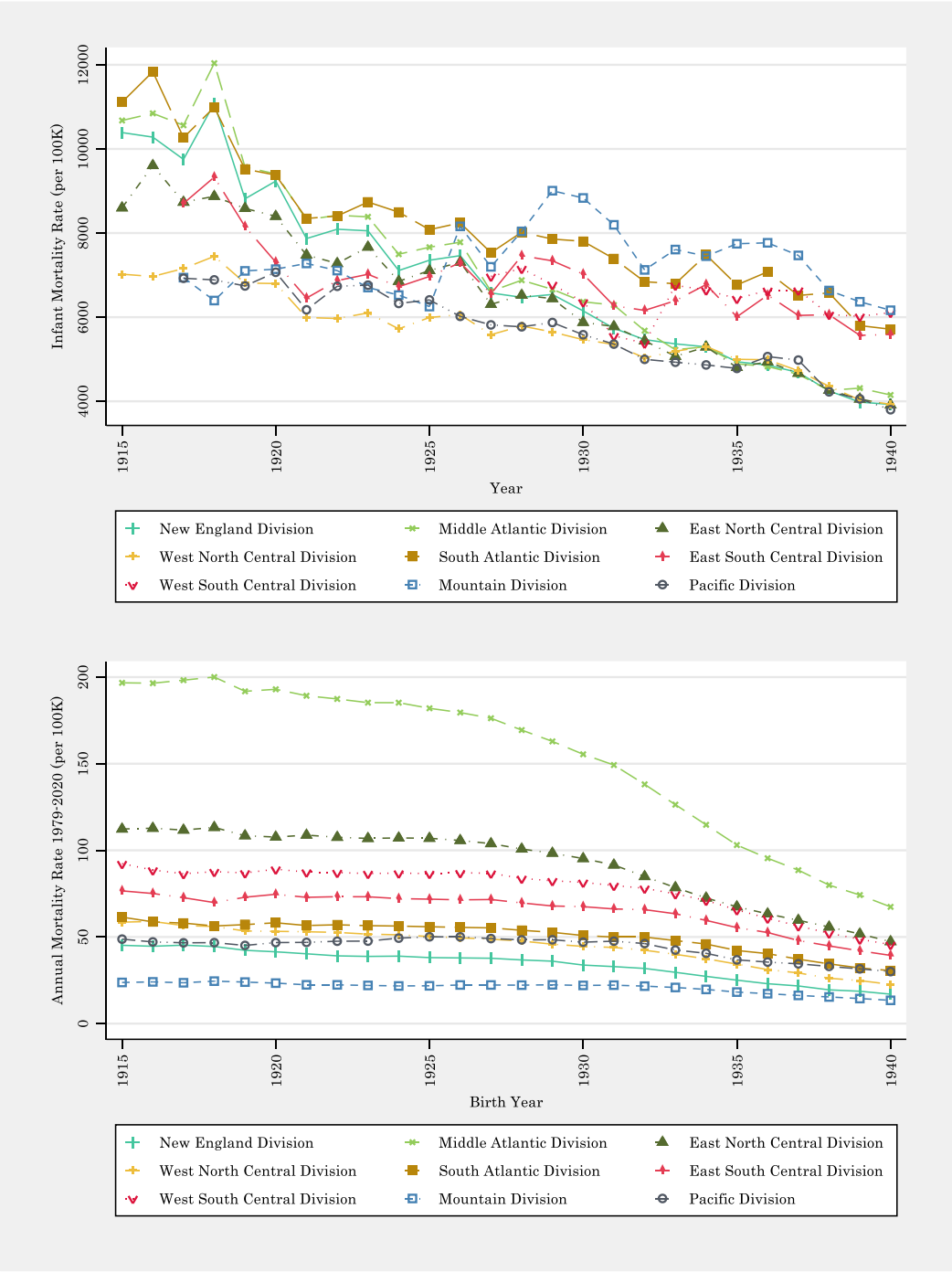



**Figure 2 - Evolution of Infant Mortality and Later-Life Mortality Rates across Birth Cohorts and Census Divisions**

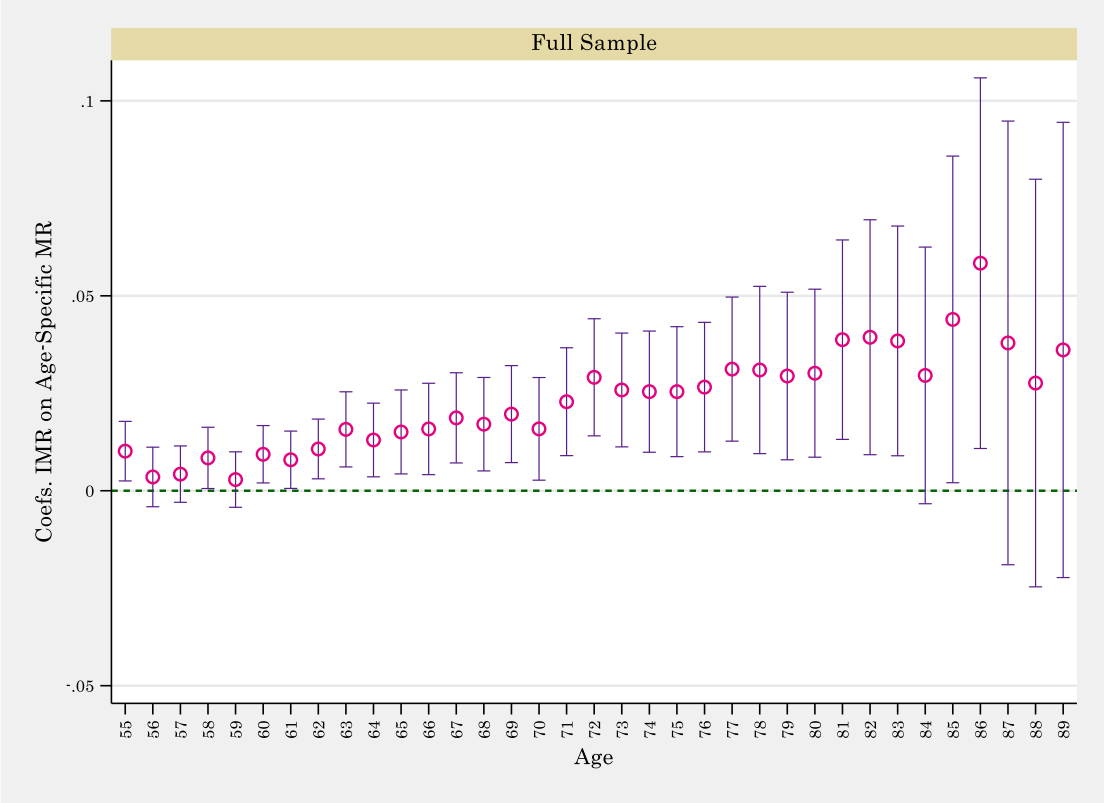



**Figure 3 - the Association between Birth-State Infant Mortality Rate and Age-Specific Mortality Rate across Different Ages**

Notes. Regression coefficients and 95% confidence intervals are illustrated. Standard errors are clustered on birth-state and birth-year. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.

## Appendix A : Data

We implement four main sample selections. First, since our main focus is exploring the latent effects of early-life environment on old-age mortality, we focus on mortality rates of ages 55 and beyond. We also restrict the cohorts to those born between 1915-1940 as these decades experienced high mortality rates as well as sharp reductions in infant mortality rates. Indeed, the infant mortality rate dropped by roughly 60% from the baseline of about 97 infant deaths per 1,000. Third, not all states participated in the death registration area and reported infant mortality rates across all years. For instance, only 10 and 24 states reported the data in the years 1915 and 1920, respectively. These numbers increased to 47 and 49 states for the years 1930 and 1940, respectively. This unbalanced panel raises the concern that the results could partly pick up on the changes in mortality due to differential longevity of later cohorts in states that joined the death registration area in later years. To address this concern, we restrict the sample to states with available infant death data from 1920 onward, i.e., focusing on 24 states. However, in Appendix D, we report the results without this selection and discuss its implications in interpreting the findings. Appendix Table A-1 reports the list of birth-states in the final sample for the years 1915 and 1920. Fourth, since SEER (2019) reports state-year-level population data of single ages up to age 89 and aggregate for all post-89 ages (the denominator values of mortality rates), we exclude deaths after age 89. After these selections, the final sample covers roughly 24.2 million deaths that occurred between 1979-2020 to birth cohorts of 1915-1940 in the age range of 55-89.

**Appendix Table A-1 - Birth States in the Final Sample for Selective Years**

| 1915 | 1920 |
|---|---|
| Connecticut | California |
| District of Columbia | Connecticut |
| Maine | District of Columbia |
| Massachusetts | Indiana |
| Michigan | Kansas |
| Minnesota | Kentucky |
| New Hampshire | Maine |
| New York | Maryland |
| Pennsylvania | Massachusetts |
| Rhode Island | Michigan |
| Vermont | Minnesota |
| | Nebraska |
| | New Hampshire |
| | New York |
| | North Carolina |
| | Ohio |
| | Oregon |
| | Pennsylvania |
| | South Carolina |
| | Utah |
| | Vermont |
| | Virginia |
| | Washington |
| | Wisconsin |

## Appendix B Functional Form

In this appendix, we examine the robustness of our results to different functional forms. First, we examine the visual correlational links between infant mortality rate and later-life mortality rate using Locally Weighted Scatterplot Smoothing (LOWESS) curves. We show the LOWESS curve for different age groups in Appendix Figure B-1. Three points emerge from this graph. First, across all age groups, exposure to higher infant mortality in early life is associated with higher later-life mortality. Second, this relationship appears to be fairly linear across different age groups, lending to our choice of functional form in equation 1. Third, these age-group-specific curves diverge from each other at higher rates of infant mortality, i.e., we observe that the slopes of curves are higher for older age groups. This implies that the latent impacts of early-life disease environment might appear larger at older ages, a fact that is later confirmed with our formal age-specific analyses.

In Appendix Table B-1, we replace the continuous measure of infant mortality rate with two dummies indicating the second and third terciles of birth-state infant mortality rate (leaving the first tercile as the reference group). Both terciles provide positive and significant coefficients, i.e., higher mortality rates due to exposure to higher infant mortality regimes. Further, the coefficient of the third tercile is roughly 70% larger than the second tercile, suggesting that the influence of early-life disease environment is higher as we move to state-years with higher infant mortality rates.

In Appendix Table B-2, we examine the sensitivity of the results to using logarithmic functional forms. In column 1, we use the log of birth-state infant mortality rate while keeping the outcome continuous. In column 2, we keep the birth-state infant mortality rate continuous and use a logarithmic outcome. In column 3, we use a log-log specification. In all cases, we observe

positive and significant coefficients. For example, a 100% percent increase in birth state infant mortality rate is associated with a 14.2% rise in later-life mortality (column 3).

As a further check, we include an interaction term of the infant mortality rate with itself in the regression model, i.e., using a polynomial regression model. If the relationship is not linear and the magnitude of effects varies significantly at different levels of infant mortality rate, then we expect to observe a positive and significant coefficient for the interaction term. The results, reported in Appendix Table B-3, only weakly support this. Although the interaction terms are positive, the fact that the interaction term is comparably small in magnitude and statistically insignificant lends support to our choice of functional form. However, the fact that the main effect is smaller in size compared to those of Table 2 and the interaction term is positive implies a small degree of nonlinearity, i.e., the impacts are larger as we move to higher levels of infant mortality rates.

**Appendix Table B-1 - Linearity of Birth State Infant Mortality and Later-Life Mortality**

| | ***Outcome: Age-Specific Mortality Rate (STD)*** | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Second Tercile of Birth-State | .01477*** | .01624*** | .02252*** |
| Infant Mortality | (.00557) | (.00584) | (.00508) |
| Third Tercile of Birth-State | .02118*** | .02333*** | .03858*** |
| Infant Mortality | (.0073) | (.00759) | (.00809) |
| Observations | 644381 | 644381 | 644381 |
| R-squared | .0283 | .06461 | .06527 |
| Mean DV | 0.000 | 0.000 | 0.000 |
| Birth-State FE | ✓ | ✓ | ✓ |
| Region-of-Birth by Birth-Year FE | ✓ | ✓ | ✓ |
| Death-State by Death-Year FE | | ✓ | ✓ |
| Birth-State Controls | | | ✓ |

Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.
*** p<0.01, ** p<0.05, * p<0.1

**Appendix Table B-2 - Birth State Infant Mortality and Later-Life Mortality: Robustness to Functional Form**

| | ***Outcomes:*** | | |
|---|---|---|---|
| | *Age-Specific Mortality Rate (STD)* | Log *Age-Specific Mortality Rate* | Log *Age-Specific Mortality Rate* |
| | (1) | (2) | (3) |
| Log Birth-State Infant | .08362*** | | .14215*** |
| Mortality Rate | (.03054) | | (.04355) |
| Birth-State Infant Mortality | | .02755*** | |
| Rate *(STD)* | | (.00954) | |
| Observations | 644381 | 644381 | 644381 |
| R-squared | .06525 | .44754 | .44755 |
| Mean DV | 0.000 | 2.783 | 2.783 |

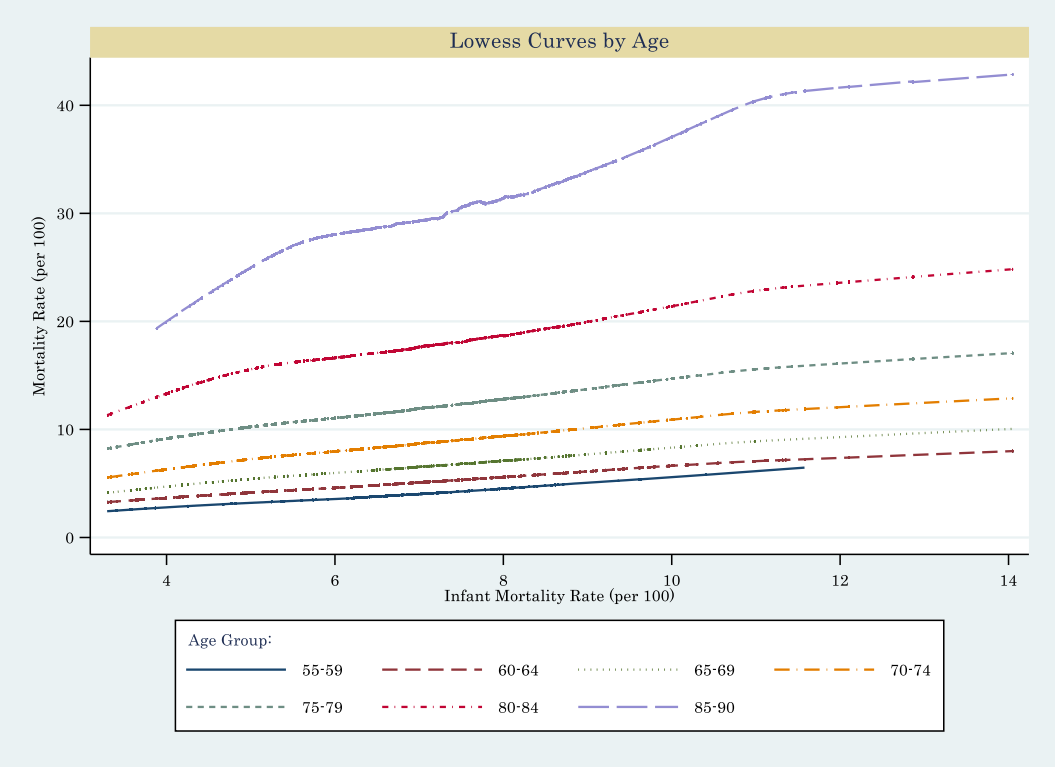
Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Variables labeled with *STD* are standardized with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.
*** p<0.01, ** p<0.05, * p<0.1

**Appendix Table B-3 - Birth State Infant Mortality and Later-Life Mortality: Examining the Nonlinearity Using a Polynomial Regression Model**

| | *Outcome: Age-Specific Mortality Rate (STD)* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Birth-State Infant Mortality Rate *(STD)* | .02605*** | .02706*** | .02023** |
| | (.00992) | (.01017) | (.00817) |
| Birth-State Infant Mortality Rate *(STD)*× Birth-State Infant Mortality Rate *(STD)* | .00292 | .0027 | .00487 |
| | (.00421) | (.00424) | (.00387) |
| Observations | 644381 | 644381 | 644381 |
| R-squared | .02834 | .06465 | .06529 |
| Mean DV | 0.000 | 0.000 | 0.000 |
| Birth-State FE | ✓ | ✓ | ✓ |
| Region-of-Birth by Birth-Year FE | ✓ | ✓ | ✓ |
| Death-State by Death-Year FE | | ✓ | ✓ |
| Birth-State Controls | | | ✓ |

Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.

*** p<0.01, ** p<0.05, * p<0.1




**Appendix Figure B-1 - Locally Weighted Scatterplot Smoothing (LOWESS) Curves of Correlation between Infant Mortality Rate and Age-Specific Mortality Rate**

## Appendix C : Heterogeneity

Early-life environment, nutrition, healthcare access, and especially exposure to the disease burden may have heterogeneous impacts on later-life outcomes across different genders and races (Almond et al., 2012; Case & Paxson, 2010). To explore such sources of heterogeneity, we replicate the main results in different subsamples based on race and gender. The results are reported in Appendix Table C-1. Contrary to previous empirical research that suggests higher impacts of exposure to disease environments for minorities and the disadvantaged population (Almond, 2006; Beach, Brown, et al., 2022), we find negative coefficients among nonwhites (columns 2). One explanation is that nonwhite subpopulations experience a considerably larger infant mortality rate which pushes selection effects to dominate the scarring effects, hence the observed reduction in their later life mortality. Further, this argument is in line with the findings of Almond et al. (2012). They examine the association between birth-state infant mortality rate and later-life maternal birth outcomes. They find that, among white mothers, higher infant mortality rates result in a higher probability of having a low-birth-weight child. In contrast, among black mothers, exposure to higher infant mortality rates is associated with a lower probability of having a low-birth-weight child. They argue that for black children, exposure to a significantly high infant mortality regime results in selection pressures outweighing the scarring impacts.

We also observe a larger impact among males and a relatively smaller coefficient for females (columns 3-4). The respective coefficient of female mortality is about 34 percent smaller in size compared with that of males, although both coefficients are statistically and economically significant. We also report the coefficients of these subsamples (female-male and white-nonwhite samples) for each age-at-death in Appendix Figure C-1 and Appendix Figure C-2.

Next, we examine the correlations across different age-cohort groups. We divide the sample into two cohorts: 1915-1927 and 1928-1940 cohorts (high and low infant mortality rate regimes in the sample). The earliest of these cohorts (born in 1915) is aged 64 at the time of starting the death window (i.e., the year 1979). The latest of these cohorts (born in 1940) are aged 80 at the time that the death window closes (i.e., the year 2020). Therefore, the age range 64-80 provides a comparable set of age-cohort sets to assess the heterogeneity by birth year. We split this age range into two similar interval groups: 64-72 and 73-80. Overall, we replicate the main findings using the following four categories: cohorts of 1915-1927 aged 64-72, cohorts of 1928-1940 aged 64-72, cohorts of 1915-1927 aged 73-80, and cohorts of 1928-1940 aged 73-80. These results are

reported in Appendix Table C-2. For the age range 64-72, both cohort groups of 1915-1927 and 1928-1940 provide quite comparable coefficients (columns 1 and 2). For the age range 73-80, on the other hand, the high infant mortality regime (i.e., born 1915-1927) provides a relatively larger coefficient compared to the low infant mortality regime (i.e., born 1928-1940).

**Appendix Table C-1 - Heterogeneity in the Correlations by Race and Gender**

| | *Outcome: Age-Specific Mortality Rate (per 100K)* | | | |
|---|---|---|---|---|
| | Whites | Nonwhites | Females | Males |
| | (1) | (2) | (3) | (4) |
| Birth-State Infant Mortality Rate | .02045*** | -.01364* | .02564*** | .03872*** |
| (per 100K) | (.00661) | (.00712) | (.00661) | (.01061) |
| Observations | 993082 | 313713 | 509070 | 552342 |
| R-squared | .04926 | .11767 | .06706 | .06946 |
| Mean DV | -0.012 | 0.120 | -0.041 | 0.038 |

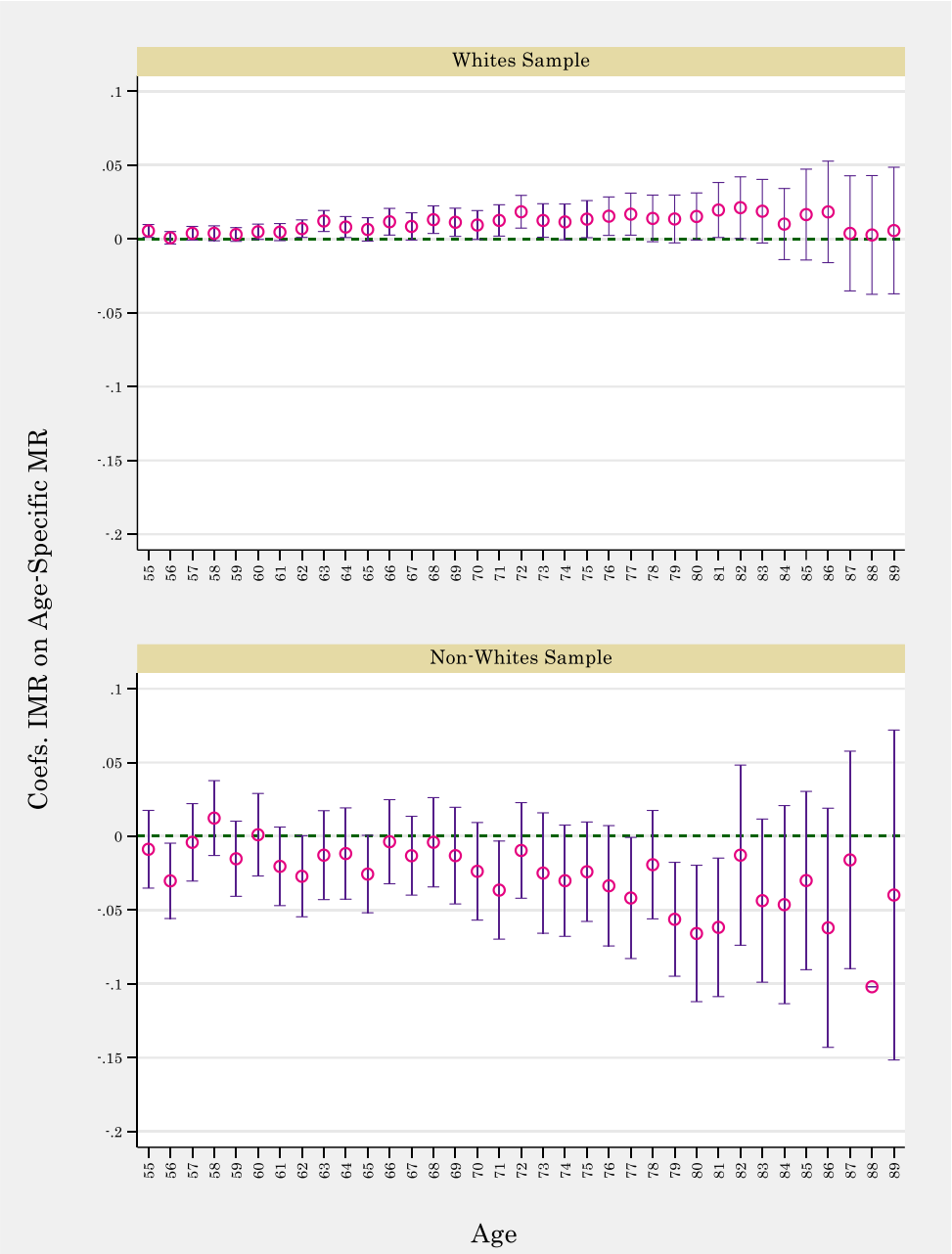
Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.

*** p<0.01, ** p<0.05, * p<0.1

**Appendix Table C-2 - Heterogeneity in Correlations of Birth State Infant Mortality and Later-Life Mortality across Ages and Cohorts**

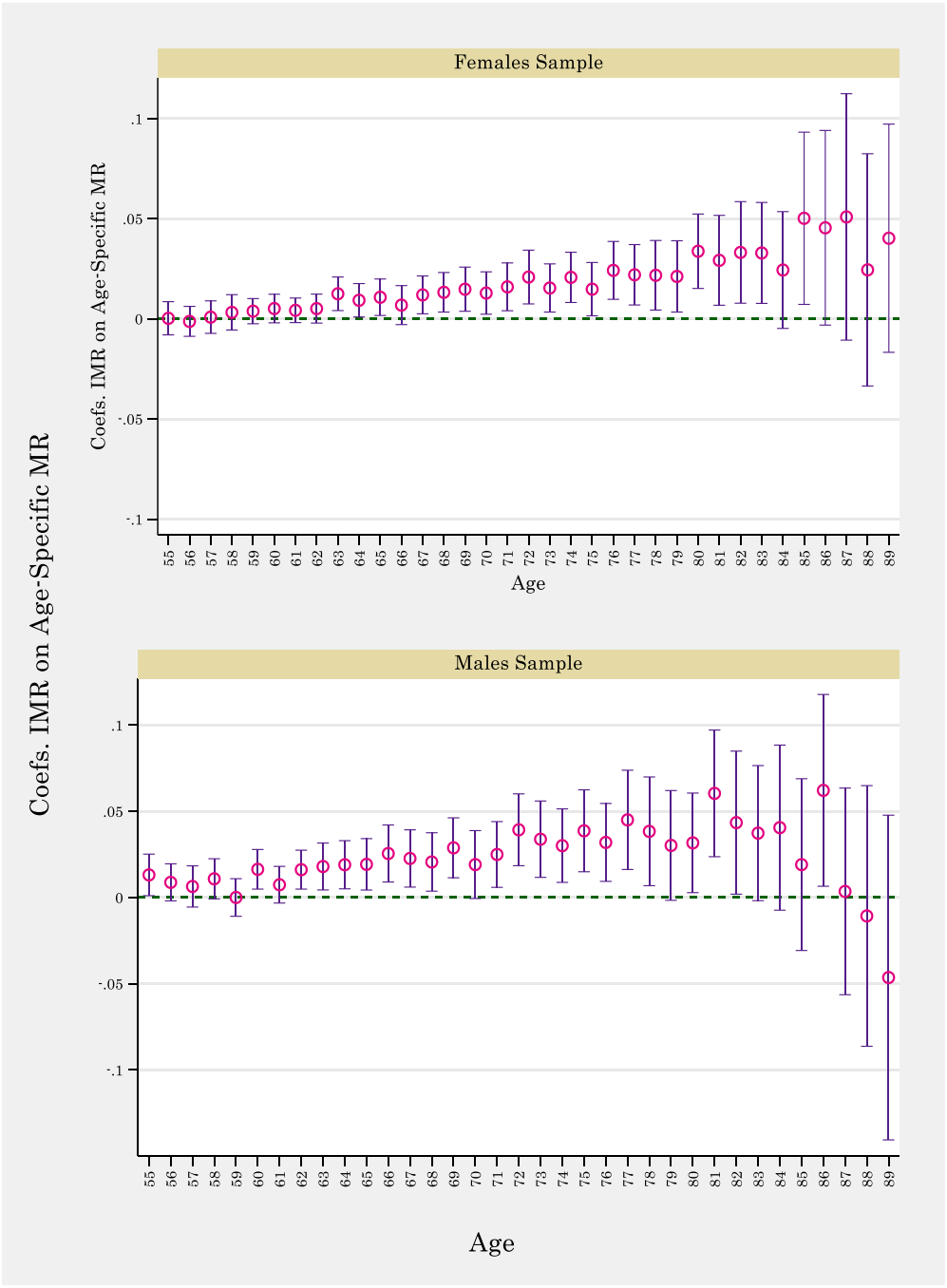
| | *Outcome: Age-Specific Mortality Rate (STD)* | | | |
|---|---|---|---|---|
| | Age 64-72; Birth Years 1915-1927 | Age 64-72; Birth Years 1928-1940 | Age 73-80; Birth Years 1915-1927 | Age 73-80; Birth Years 1928-1940 |
| | (1) | (2) | (3) | (4) |
| Birth-State Infant Mortality Rate *(STD)* | .00912*** | .00831** | .01346*** | .00895 |
| | (.0033) | (.00342) | (.00475) | (.00579) |
| Observations | 82628 | 94334 | 81332 | 93456 |
| R-squared | .07447 | .06458 | .074 | .06394 |
| Mean DV | -0.039 | -0.077 | 0.084 | 0.016 |

Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.
*** p<0.01, ** p<0.05, * p<0.1



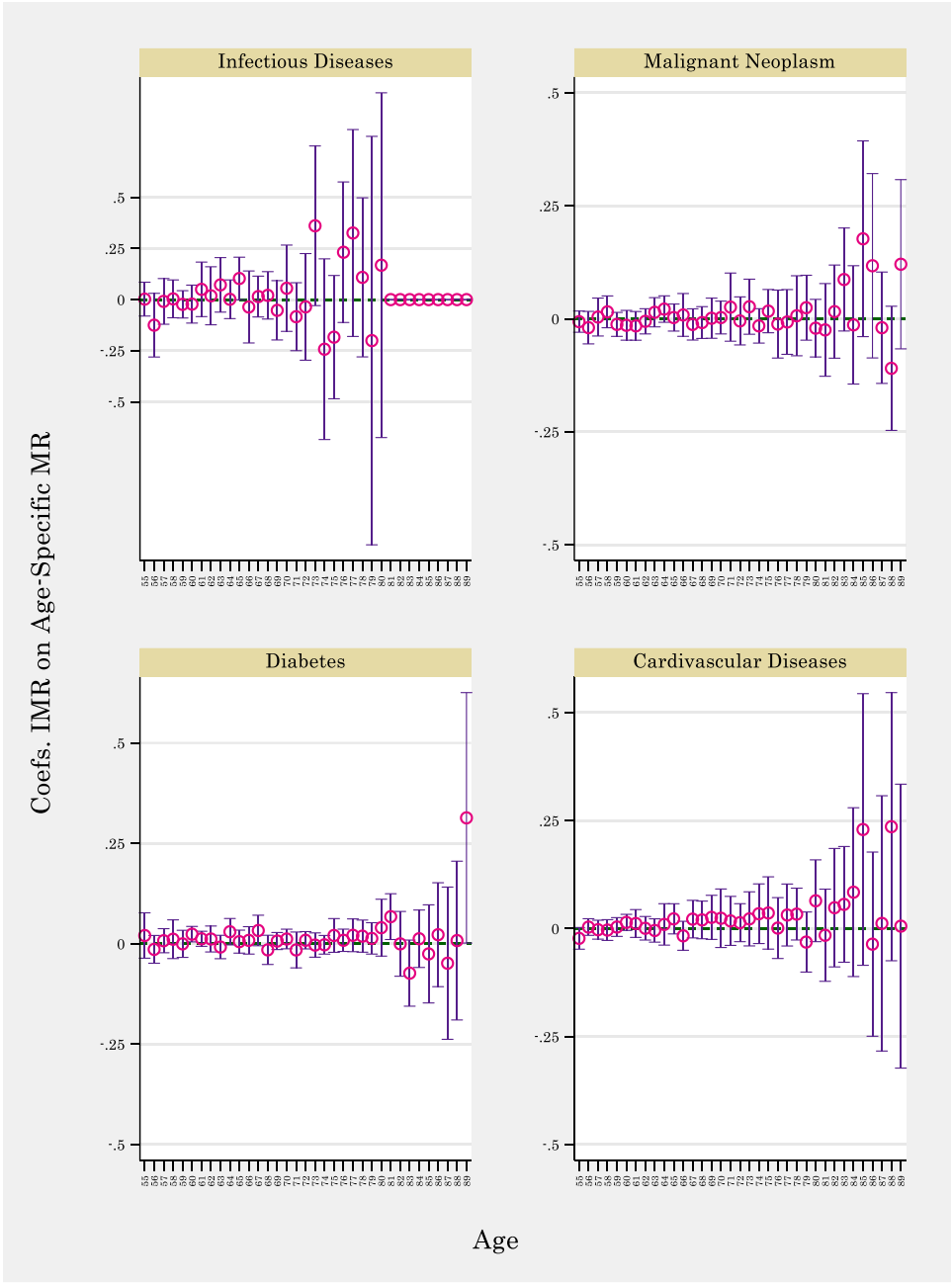

**Appendix Figure C-1 - The Association between Birth-State Infant Mortality Rate and Age-Specific Mortality Rate across Different Ages and Races**

Notes. Regression coefficients and 95% confidence intervals are illustrated. Standard errors are clustered on birth-state and birth-year. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.



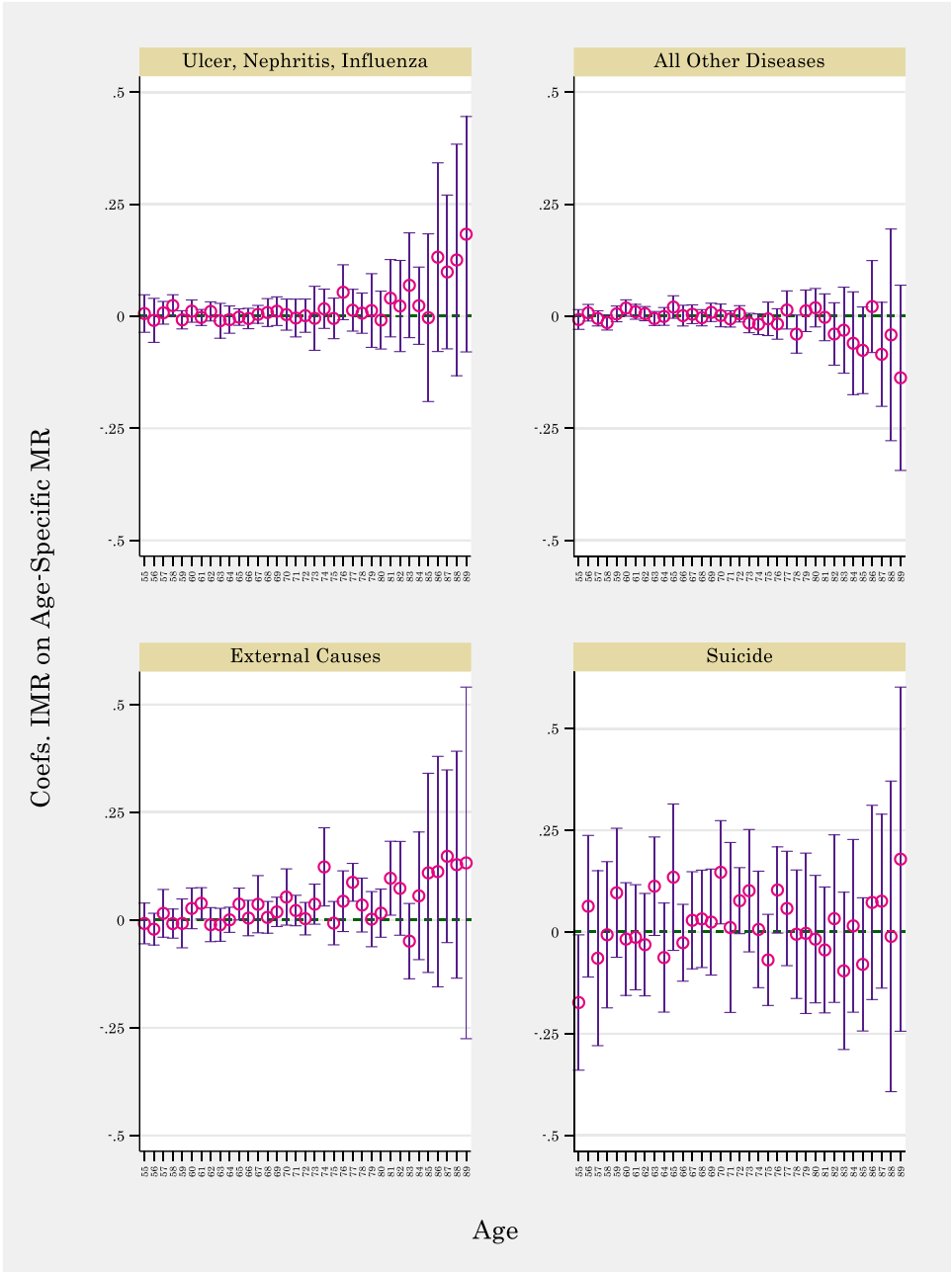

**Appendix Figure C-2 - The Association between Birth-State Infant Mortality Rate and Age-Specific Mortality Rate across Different Ages and Genders**

Notes. Regression coefficients and 95% confidence intervals are illustrated. Standard errors are clustered on birth-state and birth-year. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.




**Appendix Figure C-3 - The Association between Birth-State Infant Mortality Rate and Age-Specific Mortality Rate across Different Ages and Cause-of-Death**

Notes. Regression coefficients and 95% confidence intervals are illustrated. Standard errors are clustered on birth-state and birth-year. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.




**Appendix Figure C-4 - The Association between Birth-State Infant Mortality Rate and Age-Specific Mortality Rate across Different Ages and Cause-of-Death**

Notes. Regression coefficients and 95% confidence intervals are illustrated. Standard errors are clustered on birth-state and birth-year. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.

## Appendix D

In the main results of the paper, we restrict the sample to include birth states that consistently report infant mortality rates from the year 1920 onward. This selection eliminates the inclusion of states that joined the death registration area after 1920. In Appendix Table D-1, we show the results without this sample selection and include all states with available infant mortality rates regardless of how many years they are included in the sample. The point estimate of the full specification of column 3 suggests a smaller effect compared with that of the main results, 0.02 versus 0.03. In Appendix Table D-2, we implement a stricter sample selection and include states that have infant mortality data throughout the birth years in the final sample, i.e., 1915-1940. We observe a considerably larger coefficient size compared with that of the main results. Appendix Figure D-1 reports the associations across different ages at death for the main results as well as the two subsamples we used in Appendix Table D-1 and Appendix Table D-2.

One interpretation of these different coefficients relates to two facts that we documented in the paper. First, the impacts are relatively larger at higher levels of infant mortality. Second, the coefficients rise in magnitude at higher ages at death. The inclusion of later-joining states mechanically adds more later cohorts to the sample who are exposed to lower levels of infant mortality rates (since infant mortality drops sharply across cohorts in this period). Furthermore, these cohorts will appear at younger ages in death records compared with earlier cohorts. This is primarily because the death window closes in 2020 in our sample. Both of these elements push the coefficients downward as we include more states that joined the data in later birth years.

**Appendix Table D-1 – Replicating the Main Results without Restricting the Sample to a Balanced Panel of States**

| | *Outcome: Age-Specific Mortality Rate (STD)* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Birth-State Infant Mortality Rate (STD) | .01367** | .01346** | .01844*** |
| | (.00587) | (.00605) | (.00666) |
| Observations | 1026477 | 1026477 | 1026477 |
| R-squared | .02651 | .04694 | .04724 |
| Mean DV | 0.000 | 0.000 | 0.000 |
| Birth-State FE | ✓ | ✓ | ✓ |
| Region-of-Birth by Birth-Year FE | ✓ | ✓ | ✓ |
| Death-State by Death-Year FE | | ✓ | ✓ |
| Birth-State Controls | | | ✓ |

Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.

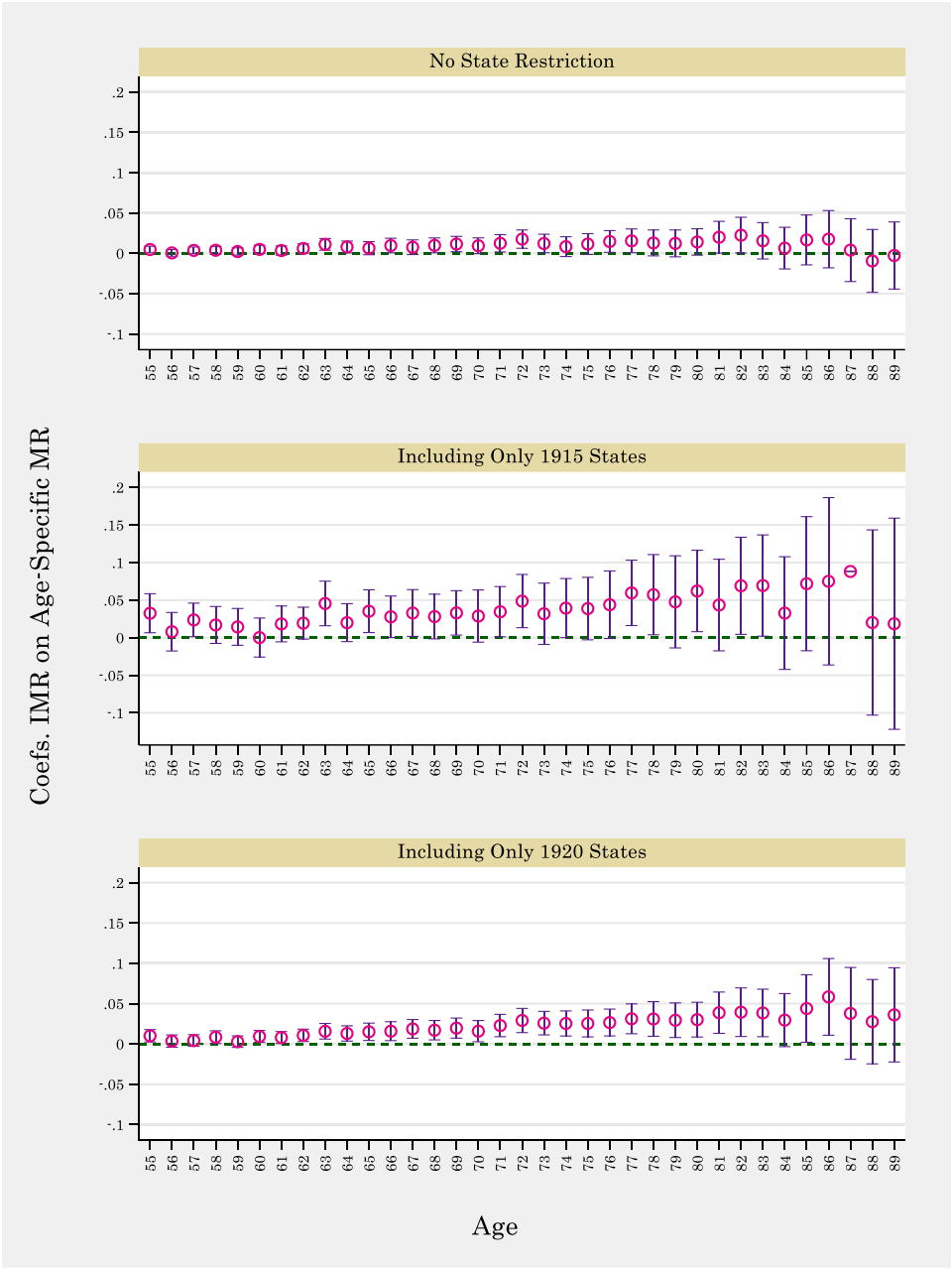
*** p<0.01, ** p<0.05, * p<0.1

**Appendix Table D-2 - Replicating the Main Results with Restricting the Sample to a Balanced Panel of 10 States Reporting Infant Death across All Years of 1915-1940**

| | *Outcome: Age-Specific Mortality Rate (STD)* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Birth-State Infant Mortality Rate (STD) | .054*** | .05557*** | .05863*** |
| | (.01967) | (.01983) | (.01792) |
| Observations | 301077 | 301077 | 301077 |
| R-squared | .04261 | .122 | .12385 |
| Mean DV | -0.000 | -0.000 | -0.000 |
| Birth-State FE | ✓ | ✓ | ✓ |
| Region-of-Birth by Birth-Year FE | ✓ | ✓ | ✓ |
| Death-State by Death-Year FE | | ✓ | ✓ |
| Birth-State Controls | | | ✓ |

Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.

*** p<0.01, ** p<0.05, * p<0.1




**Appendix Figure D-1 - the Association between Birth-State Infant Mortality Rate and Age-Specific Mortality Rate across Different Sample Restrictions**

Notes. Regression coefficients and 95% confidence intervals are illustrated. Standard errors are clustered on birth-state and birth-year. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. All regressions include birth-state fixed effects, birth-year fixed effects interacted by region-of-birth fixed effects, death-state by death-year fixed effects, and birth-state controls. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.

## Appendix E

The denominator of mortality rates in the paper is calculated based on current state-year population counts. The reason is that our data source regarding population estimates does not report population based on the level of mortality estimates in the numerator, i.e., birth state, current state, birth year, current year, and age. While we acknowledge this data limitation in our mortality estimates, we further examine the robustness of the results using an alternative data source for the denominator. Specifically, we use population estimates based on random samples of decennial censuses combined with the American Community Survey data. The disadvantage of this data source is that it does not cover inter-decennial years 1981-1989 and 1991-1999. Further, the estimates are based on random samples rather than the full population. The main advantage, however, is that we can estimate the population at the numerator level (i.e., birth state, current state, birth year, current year, and age). Therefore, for the years 1980, 1990, and 2000-2020, we can calculate mortality rates using numerators and denominators at the same level of aggregation. The results are reported in Appendix Table E-1. The fully parameterized regression of column 3 is almost identical to the main results.

**Appendix Table E-1 - Replicating the Main Results Using Census-Based Denominator**

| | *Outcome: Age-Specific Mortality Rate (STD)* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Birth-State Infant Mortality Rate *(STD)* | .02231* | .01778 | .03114*** |
| | (.01271) | (.0116) | (.01141) |
| Observations | 234812 | 234812 | 234812 |
| R-squared | .05413 | .15519 | .15549 |
| Mean DV | -0.000 | -0.000 | -0.000 |
| Birth-State FE | ✓ | ✓ | ✓ |
| Region-of-Birth by Birth-Year FE | ✓ | ✓ | ✓ |
| Death-State by Death-Year FE | | ✓ | ✓ |
| Birth-State Controls | | | ✓ |

Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.

*** p<0.01, ** p<0.05, * p<0.1

## Appendix F

Appendix Table F-1 presents lead–lag estimates examining how exposure to birth-state infant mortality around the year of birth is associated with later-life age-specific mortality. The coefficients capture the relationship between infant mortality rates in the birth state from three years before birth through five years after birth. The estimates show little evidence of pre-trends: infant mortality rates in the years preceding birth are small and statistically insignificant, supporting the validity of the identification strategy. In contrast, infant mortality in the year of birth is positively and significantly associated with later-life mortality, with a one-standard-deviation increase corresponding to a 2.3% higher mortality rate in old age. The effect persists, though attenuated, in the year immediately following birth and fades in subsequent years. These patterns suggest that in utero and the immediate postnatal period represent the most sensitive windows during which exposure to adverse disease environments affects later-life mortality.

**Appendix Table F-1 - Lead–Lag Estimates of Birth-State Infant Mortality and Later-Life Mortality**

| | ***Outcome: Age-Specific Mortality Rate (STD)*** | |
|---|---|---|
| | (1) | (2) |
| *Birth-State Infant Mortality Rate (STD) at:* | | |
| Year of Birth - 3 | .00491 | .00496 |
| | (.0085) | (.00864) |
| Year of Birth - 2 | .01228 | .01258 |
| | (.00842) | (.00846) |
| Year of Birth – 1 | .01236 | .01097 |
| | (.00836) | (.00844) |
| Year of Birth | .02132** | .02256** |
| | (.00934) | (.00932) |
| Year of Birth + 1 | .01786* | .01777* |
| | (.00987) | (.00986) |
| Year of Birth + 2 | .00564 | .00674 |
| | (.00843) | (.00851) |
| Year of Birth + 3 | .00347 | .00399 |
| | (.00909) | (.00918) |
| Year of Birth + 4 | .00789 | .00682 |
| | (.00842) | (.0085) |
| Year of Birth + 5 | .00173 | .00184 |
| | (.00879) | (.00883) |
| Observations | 591104 | 591104 |
| R-squared | .02865 | .06639 |
| Mean DV | -0.005 | -0.005 |
| Birth-State FE | ✓ | ✓ |
| Region-of-Birth by Birth-Year FE | ✓ | ✓ |
| Death-State by Death-Year FE | | ✓ |
| Birth-State Controls | | ✓ |

Notes. Standard errors, clustered on birth-state and birth-year, are in parentheses. Both independent variable and dependent variable are standardized (STD) with respect to their sample's mean and standard deviation. Controls include birth-state share of white-collar workers, share of farmers, share of other occupations, socioeconomic index, share of literate individuals, share of midwives, and share of doctors. Regressions are weighted using the average number of state-level age-level population in the 1980 census.
*** p<0.01, ** p<0.05, * p<0.1