Long Run Consequences of Exposure to Natural Disasters*

Krzysztof Karbownik[†] Anthony Wray[‡]

October 30, 2015

Abstract

We study the long-run effects of in-utero exposure to hurricanes in the U.S. during the late-nineteenth century. A difference-in-differences regression identifies the effect of hurricane exposure by comparing exposed individuals to those born in the same location before and after the storm, in addition to individuals born in neighboring locations concurrent to the storm. We find that *in utero* exposure to a hurricane reduces educational attainment by a quarter of a year and the probability of high school completion by 20 percent. Sibling-fixed-effects regression results suggest that in-utero exposure is associated with a 1.5-year (12 percent) reduction in life expectancy at age sixty-five.

Latest version here.

*Wray is indebted to his dissertation committee members Joel Mokyr (chair), Joseph Ferrie and David Dranove for encouragement and guidance. We also thank seminar participants at the Northwestern University Applied Micro Lunch, and conference participants at the Early Life Conference at the University of Michigan (2014), the Canadian Economic Association Annual Meeting (2014) for helpful comments and suggestions. Wray appreciates the financial support from the Northwestern University Economics Department's Eisner Fund and Center for Economic History, and the Economic History Association's Sokoloff Fellowship. All errors of omission are our own.

[†]Institute for Policy Research, Northwestern University; Email: krzysztof.karbownik@northwestern.edu.

[‡]Hitotsubashi Institute for Advanced Study, Hitotsubashi University; Email: anthony.wray@r.hit-u.ac.jp.

1 Introduction

The immediate loss of life and destruction of property caused by natural disasters is well-documented, but there is scant evidence on the long-run consequences for survivors affected by natural disasters. Climate change and the growth of coastal populations have brought an increasing share of the U.S. population into geographical areas at risk of natural disasters such as hurricanes and floods.[1] The growing human and economic toll of these extreme weather events makes it imperative to understand the long run health and socioeconomic costs of exposure to such environmental shocks.

In this project, we study the long-run effects of in utero and early childhood exposure to hurricanes in the U.S. during the late-nineteenth century. A large body of research over the past two decades has studied the medium- and long-term consequences of early-life circumstances, particularly those related to health (Barker, 1995). A consensus has emerged among researchers that adverse shocks to in utero or early childhood health lead to worse birth outcomes, lower education and wages, and can even lead to the intergenerational transmission of the scarring effect from shocks.[2] A growing subset of this literature has studied how environmental or economic shocks can affect children, either biologically or economically, through a disruption to the availability of resources.[3] These studies suggest that there are potentially very large returns over the long-run to improving in utero or childhood circumstances.

Our project examines whether the effects of in utero and early childhood exposure to hurricanes persist late into life, a time horizon far longer than has been documented by the existing literature. Recent work has shown that in utero exposure to hurricanes affects a wide variety of birth outcomes, such as prematurity and abnormal conditions of the newborn

[1]While the population at risk has increased, Table 1 shows that the frequency of all hurricanes and the frequency of major hurricanes (Category 3 or higher) has remained relatively constant between the beginning of the twentieth century and the first decade of the twenty-first century.

[2]See Almond and Currie (2011) for a review of this literature.

[3]Environmental shocks studied by the existing literature include earthquakes (Torche, 2011), hurricanes (Currie and Rossin-Slater, 2013), radiation (Almond *et al.*, 2009)), rainfall (Maccini and Yang, 2009), and wildfires (Jayachandran, 2009).

(Currie and Rossin-Slater, 2013; Simeonova, 2011; Sotomayor, 2012), as well as rates of infant mortality (Antilla-Hughes and Hsiang, 2013). Other studies find effects on medium-run schooling measures (Crittenden Fuller, 2012; Deuchert and Felfe, 2013). However, evidence that any type of environmental or economic shocks has a persistent effect on later-life outcomes is limited and mixed.[4]

In addition to whether the effect of exposure to hurricanes persists into adulthood and old age, it is equally unclear whether the underlying mechanism would be biological, via maternal (child) stress during pregnancy (infancy), or economic, through loss of personal income and property, or disruptions to resource availability and infrastructure. Whether in utero exposure alone, or early-childhood exposure as well, have long-run effects will provide some insight into the possible mechanisms. Moreover, large and persistent effects of early-childhood exposure to shocks would entail a much larger scope for policy intervention than the effects of in utero exposure, on which previous literature has focused.

In addition to studying medium-run effects of exposure to natural disasters on educational outcomes, we conduct the first analysis of the long-term effects on longevity. To do so, we focus on in utero and early-childhood exposure to hurricanes and conduct two separate empirical exercises. First, we identify individuals from the 1886 to 1897 birth cohorts residing in 100-km bands around areas exposed to hurricanes in the late-19th century and link them to educational outcomes in the 1940 U.S. Population Census. Second, we link individuals from the 1896 to 1900 birth cohorts, similarly residing in the hurricane exposure zone, to the Social Security Administration's Death Master File, in which we observe longevity. The use of historical data permits us to follow the entire life-cycle of individuals, up to and including death. The primary advantage of focusing on hurricanes is that while they cover a large area

[4] Cutler *et al.* (2007) do not find any significant effects of in utero exposure to poor economic conditions during the Dust Bowl on mortality and later-life health. Lindeboom *et al.* (2010) and van den Berg *et al.* (2011) study the effect of business cycle conditions at birth and in utero exposure to famine, respectively, on life-cycle and later-life outcomes, such as mortality, using proportional hazard models. A recent study (Aizer *et al.*, 2013) examines the effects of a mothers pension receipt using similar measures of life-cycle outcomes as this project. As our analysis makes comparisons within families, through the use of mother-fixed effects estimation, we believe that we can produce more convincing and stronger causal estimates than previous studies.

and cause a large amount of destruction, the typical death toll has historically been small relative to other exogenous shocks, such as the 1918 flu pandemic (Almond, 2006). This fact ensures that our analysis of adjacent birth cohorts around the occurrence of a storm is less likely to suffer from selective mortality. In addition, hurricanes return to the same location infrequently enough that individuals were unlikely to have perfectly anticipated future storms, thus reducing the possibility of residential sorting on disaster risk prior to storms.

The remainder of the paper is structured as follows: Section 2 describes the data sources, Section 3 explains how the data sources are linked together, Section 4 provides details on geolocating individuals and defining exposure to hurricanes, Section 5 shows the empirical specifications and the threats to identification, Section 6 presents the main regression results, Section 7 provides robustness exercises, and Section 8 concludes.

2 Data sources

2.1 Hurricane Tracks

We reconstruct hurricane paths using the HURDAT2 database of best-track estimates for Atlantic hurricanes produced by the National Hurricane Center (NHC) at the National Oceanic and Atmospheric Administration (NOAA).[5] The HURDAT2 database contains all storms that occurred between 1851 and 2012 in the North Atlantic basin, and matched or exceeded the intensity of tropical depressions. It contains a standardized set of latitude and longitude coordinates for the location of a storm every six hours, which are estimated from storm observations in historical sources. We use GIS software to join the coordinates into storm tracks.

Table 2 lists the nine hurricanes with a minimum of Category 1 strength that made U.S.

[5] http://www.aoml.noaa.gov/hrd/hurdat/easyread-2012.html. For more details on the re-analysis project that updated the original HURDATA database in 2012, see http://www.aoml.noaa.gov/hrd/hurdat/Data_Storm.html.

landfall between 1896 and 1899, and provides estimates of wind speed and damage costs. We construct indicator variables for hurricane exposure at the county-by-conception-week level and assume a standard 39-week gestation length. While the 1900 census only reports the month and year of birth, the exact date of birth reported in the World War I Draft Registration Cards (WWI records) and the Death Master File (DMF), discussed below in Section 2.3, allows us to identify the in utero period at a weekly frequency.

2.2 Historical U.S. Population Censuses

We obtain information on the individuals that comprise our estimation sample from the complete count files of historical U.S. Population Censuses. We begin by extracting the records of all males from the 1886 to 1900 birth cohorts in the 1900 census, from which we obtain background characteristics for the cohorts of interest. We restrict attention to males because name changes at marriage prevent the matching of females to long-run outcomes. The records are grouped by household, allowing us to identify siblings and construct additional individual-level variables such as birth order and mothers age at birth. A crucial feature of the 1900 census is that it is the only historical U.S. census that reported an individuals month and year of birth, whereas other censuses only recorded age in years on the last birthday. The additional detail on the timing of birth improves the precision of matches to the database of World War I Draft Registration Cards and to the Death Master File. We turn to the 1940 census to obtain the educational outcomes – years of education and grade completion indicators – for the 1886 to 1897 birth cohorts.[6]

2.3 U.S. World War I Draft Registration Cards

Determining whether an individual was exposed to a hurricane while in utero requires precise information on the date of birth, as well as the place of birth. However, the U.S.

[6] The analysis of educational outcomes is restricted to individuals born as late as the 1897 birth cohort because individuals born later were too young to register for the WWI draft. See Section 2.3 below for more details.

Population Census records alone are insufficient as they only report place of birth at the state level. We supplement the census records with the U.S. World War I Registration Cards of approximately 2 million individuals from the 1886 through 1897 birth cohorts born in one of the following states: Alabama, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Texas, and Virginia. These records contain an individuals name, exact birth date, and place of birth at either the county, city or township level. In the absence of the detailed place of birth information, it would not be possible to identify the in utero hurricane exposure status of individuals who had migrated by the time they had been enumerated in the 1900 census.

Given that the Selective Service Act of 1917 required all males residing in the U.S. between the ages of 18 and 45 to register for the draft, the records should include the near universe of males from the 1886 to 1897 birth cohorts.[7] One caveat is that our data is restricted to the First and Second Registrations, because the cards from the Third Registration did not report place of birth. While we can identify everyone from the 1886 through 1896 birth cohorts, who were covered by the First and Second Registrations, we will only capture individuals born before 24 August 1897, as younger individuals were covered by the Third Registration.

2.4 Social Security Administration's Death Master File

The public release version of the Social Security Administrations Death Master File (DMF) includes an individuals name, exact dates of birth and death, and Social Security

[7]On 18 May 1917, six weeks after the U.S. declared war on Germany, the U.S. government passed the Selective Service Act, which required all males living in the U.S. between the ages of 18 and 45, regardless of citizenship status, to register for the draft. Three separate registrations were held for the World War I draft. The First Registration took place on 5 June 1917, and covered men born between 6 June 1886 and 5 June 1896 (those aged 21 to 31 at the time). The Second Registration, which took place on 5 June 1918, covered males born between 6 June 1896 and 5 June 1897 (those who had turned 21 since the previous registration), as well as those covered by the First Registration who had not already registered. A supplemental registration was held on 24 August 1918 to cover males who had turned 21 since 5 June 1918. The Third Registration was held on 12 September 1918 and covered males aged 18 to 21 (born between 11 September 1897 and 12 September 1900), and aged 31 to 45 (born between 11 September 1872 and 12 September 1887). For more details, see http://www.search.ancestrylibrary.com/DB.aspx?dbid=6482.

Number. We obtain measures of longevity by linking individuals to the DMF, which contains all deceased individuals who received Social Security benefits during their lifetimes. Hill and Rosenwaike (2001) show that the completeness of the DMF was severely limited for all age groups until the mid-1960s. They find that the coverage of deaths among individuals aged 65 or older exceeded 93 percent for most years between 1973 and 1997, with a low point of 80-percent coverage in 1987. Given that the cohorts of interest in our preliminary analysis turned 65 years old between 1960 and 1964, we may observe a sample of individuals in the DMF records selected on the basis of their interaction with the Social Security system.[8] The DMF includes some individuals who passed away before the age of 65 if they had received Social Security disability benefits. Given that this is a highly selected sample of individuals, we restrict attention to individuals who survived until age 65.

3 Data linkage

3.1 Linking U.S. census data to WWI records

Figure 5 provides a visual representation of the data sources and illustrates the linking procedure used in the analysis of the educational outcomes. We match the study cohorts from the 1900 U.S. Population Census to the World War I Draft Registration using six linking variables: first name, middle initial, surname, state of birth, and month and year of birth. The 1940 census also contains state of birth, but it does not report month of birth and so we must drop that variable in the link from the 1900 to 1940 census. The first stage of the linking procedure involves taking as a potential match any pairs of observations that meet at least one of the following criteria:

(i) First letter of given name matches AND surname SOUNDEX code matches AND year of birth is within 2 years.

[8] In future work, we will estimate the effect of hurricane exposure on life expectancy conditional on survival to age 75, and consider later birth cohorts as well as alternative sources of data for outcome measures.

(ii) First letter of surname matches AND given name SOUNDEX code matches AND year of birth is within 2 years.

We repeat (i) and (ii) by replacing the SOUNDEX code with the Atack-Bateman implementation of the NYSIIS algorithm, or the first three consonants of the name, and take the union of matched observations across the criteria. In the second stage of the linking procedure, we exclude observations that have different non-missing middle initials, and rank the remaining observations based on match quality. We impose penalties for the Levenshtein distances of the first name and surname weighted by the length of each string (allowing a maximum weighted distance of 0.4 for each variable), and the absolute deviation in year of birth between the sources. We prioritize observations with a closer match on age, then accept observations that uniquely minimize the length-weighted Levenshtein distance averaged over the first name and surname. In the case of linking the 1900 census to the WWI records, we also require an exact match on month and year of birth. Table 8 shows the matching rates and samples sizes.

3.2 Linking the 1900 U.S. census to the Death Master File

The analysis of the longevity outcomes is based on a sample linking the 1900 U.S. population census to the Social Security Administration's Death Master File. We make use of four linking variables: first name, surname, year of birth and month of birth. Noticeably missing from the linking criteria is place of birth, which is not available in the DMF and only recorded at the state level, but this is offset by improved precision of matches from knowing the month of birth. The matching procedure follows the algorithm described in Section 3.1.

Given that the 1900 census only records birth place at the state level, we assume that the county of birth is the same as the county of residence in 1900, and we exclude individuals residing outside their state of birth in 1900. To reduce measurement error caused by households that moved across county lines, we restrict attention to the 1895 to 1899 birth cohorts for the analysis of the longevity outcomes.

4 Geocoding birth locations and hurricane tracks

4.1 Geocoding Birth Location

We obtain the latitude and longitude coordinates for an individuals birth location by matching the place of birth strings in the World War I Draft Registration Cards to the universe of U.S. place names in the Geographic Names Information System (GNIS) file from the USGS. The database includes historical features that no longer exist on the landscape. We matched 84 percent of the birth locations in the WWI records to the latitude and longitude coordinates for a city or town. Figure 4 displays the population weighted locations matched to the GNIS file. Next, we linked the remaining records to city and county names in the 1900 census, and assigned them to the coordinates of the county centroid. These matches account for an additional 10 percent of the WWI records.

4.2 Ambiguity in birth locations

A difficulty with geolocating the birth places listed in the World War I draft registration cards is that some place names may refer to a county name as well as the name of a city in another county. Such ambiguity can introduce significant measurement error to the determination of hurricane exposure status. To remedy this problem, we make use of information on the residential locations of individuals enumerated in the 1900 U.S. census. Recall that the WWI records were matched to the 1900 census using only the state of birth to identify birth location. For each unique birth-place location string in the WWI records that appeared at least 10 times in WWI-1900 matched sample, we calculated the modal county of residence in 1900. If at least 50 percent of the individuals resided in the modal county, then we took that location as the probable county of birth. In other words, we excluded locations for which the implied inter-county migration rates were implausibly large.

4.3 Definitions of hurricane exposure

We restrict our estimation samples to individuals born between 1885 and 1897, in the case of the educational outcomes, or between 1896 and 1900, in the case of longevity outcomes, in locations within 100 kilometers of a hurricane track that made continental U.S. landfall. We use two separate criteria to determine the set of observations to include in the analysis, depending on the precision of available information on birth location. First, in the analysis of educational outcomes, the birth location is known at the city or county level from the WWI records. Individuals are considered *at risk* of hurricane exposure if the coordinates of the birth location overlapped a 100-kilometer radii buffer zone around a hurricane-strength storm track. When birth location is known only at the county level, the coordinates of the county centroid are used to proxy for distance from the hurricane track. Second, in the analysis of longevity outcomes, we proxy for birth location using the county of residence in the 1900 census. Individuals are considered *at risk* of hurricane exposure if the county of residence at least partially overlapped a 100-kilometer radii buffer zone around a storm track.[9]

Figure 1 displays a map of 1900 county boundaries for the Southern U.S. states from NHGIS. It shows the counties in our sample and highlights the counties that overlap 40-kilometer-radii buffer zones around a storm track. A birth location-week cell is considered exposed to a hurricane if the location overlaps a 40-kilometer-radii buffer zone around a hurricane track for that week. We assign a vector of indicator variables to each location-by-conception-week cell, such that each element of the vector captures hurricane exposure for a given week relative to the week of conception. These elements are aggregated to form the variables used in the models. In robustness checks, we apply varying definitions for locations exposed to a hurricane: 30, 40 or 50 km radii buffer zones around the hurricane tracks. Figure 2 shows the population weighted locations for all observations in the WWI records

[9] We define counties according to the 1900 boundaries contained in shapefiles provided by the National Historical GIS (NHGIS).

within the 100 km buffer zones, while Figure 3 shows the same for the matched sample.

5 Empirical Specifications

To examine effects on educational outcomes, we estimate a difference-in-differences (D-in-D) regression in which the effect of hurricane exposure is identified by comparing exposed individuals to those individuals born in the same county before and after the storm, as well as to those individuals born in neighboring counties concurrent to the storm. Specifically, we estimate:

$$Y_i = \beta_0 + \beta_1(\textit{In-utero}_i) + \sum_{n=1}^{3} \beta_2^n (\textit{Age } n)_i + \pi X_i + \alpha_c + \gamma_y + \delta_m + \epsilon_i \tag{1}$$

where Y_i denotes one of the outcome variables (years of completed education, probability of high school completion, or the number of years lived beyond the age of 65) for individual i, $(\textit{In-utero}_i)$ represents the effects of in utero exposure to a hurricane, $(\textit{Age } n)_i$ represents the separate effects of exposure at age 1 through age 3, and X_i contains an indicator for conception during the one-year period after the storm, as well as indicator variables for race, birth order, maternal age at birth, married mother, and whether the mother had reported a death of a child. We include county of birth (α_c), conception year (γ_y), and conception month (δ_m) fixed effects. Standard errors in this equation are clustered at the county level. We define exposure as having a birth location or county of residence located within 40 kilometers of a storm track during the relevant period. We also exclude a "donut hole" region from the estimation sample consisting of counties 40 to 70 kilometers away from the hurricane track, to account for uncertainty in the size of the area exposed to the storm.

Given that the D-in-D regression does not control for unobserved factors common to a household that may be correlated with long-run outcomes and the timing of conception relative to a hurricane, we also estimate a mother-fixed-effects regression:

$$Y_i = \beta_0 + \beta_1(\textit{In-utero}_i) + \sum_{n=1}^{3} \beta_2^n(\textit{Age } n)_i + \pi X_i + \alpha_j + \gamma_t + \delta_m + \epsilon_i \qquad (2)$$

where, in comparison to the D-in-D regression, we replace the county fixed effects with a mother fixed effect α_j, and X_i includes controls for mother's age at birth and the birth order among siblings of both genders. Standard errors in this equation are clustered at the mother level. Due to insufficient sample size after linking together three sources, we only estimate the sibling fixed effects model for the longevity outcomes. In addition to the continuous measure of the number of years lived beyond the age of 65, we also define a binary dependent variable equal to one in the case of the sibling that died at a younger age.

5.1 Threats to identification

Measurement error in the assignment of hurricane exposure status poses a potential threat to our identification strategies. Although the D-in-D specification controls for the effect of county specific shocks, it will not pick up local variation in storm damage and relief efforts. The resulting measurement error will bias estimates of the effect of hurricane exposure in the D-in-D specification toward zero. Furthermore, the definition of hurricane exposure as a binary variable may not capture gradations in the effect of exposure by distance from the hurricane path. To the extent that individuals located more than 40 km from the hurricane path are negatively affected by the storm in some way, but are assigned unexposed status, our D-in-D estimates will be biased downward. These problems are not a concern when we use the mother-fixed-effects specification.

Another limitation of our empirical specifications is that neither deals with the possibility of endogenous migration in response to a disaster, which may change the composition of women giving birth in the affected area following the disaster. Census records contain state of birth, a level of geographic precision which is insufficient to identify whether the individual was exposed to a hurricane around the time of birth. Thus, we lack the information to identify intra-state migrants and must assign storm exposure to individuals based on their county of residence when enumerated in the 1900 census. If wealthier families possess the resources to migrate, then we may observe worse births outcomes among the greater proportion of disadvantaged mothers that remain. However, the effect could go in either direction. Boustan *et al.* (2012) used linked US census data from the 1920s and 1930s to show that existing residents and in-migrants were less likely to live in State Economic Areas (SEA) that had recently experienced a tornado, but more likely to reside in SEAs that had been flooded.

Another limitation of using historical data is that, unlike contemporary studies, we do not observe gestation length. Our assumption that all pregnancies lasted exactly 39 weeks (or 9 months) raises the possibility that a premature infant will be recorded as exposed during the third trimester, which induces measurement error in the indicator of exposure during the first and second trimesters.[10] In addition to the issues with data quality, a mechanical correlation exists between longer gestation length and the probability of exposure to a hurricane at any point of the pregnancy (Currie and Rossin-Slater, 2013). As these factors introduce false positives to the set of exposed individuals, our coefficients will be biased towards zero, and must be interpreted as lower bound estimates.

6 Results

6.1 Education results

We begin by presenting results from estimating Equation 1 with four different educational outcomes as the dependent variables, and display the results from the main specification in Table 4. In column (1) the dependent variable is an individual's years of completed schooling, while in columns (2) to (4) the dependent variables are indicators for completion

[10] We will test the sensitivity of the assumption by performing robustness checks in which all pregnancies are assigned a gestational age of 38 to 41 weeks.

of primary school (grade 6), secondary school (grade 12), and college (4 years of post-secondary), respectively. We find that in-utero exposure to a hurricane reduces educational attainment by 0.24 years or 0.06 standard deviations. We also show that hurricane exposure is associated with a 3.8 percentage point decline in the probability of high school completion, which is a 19.7 percent effect relative to the mean completion rate. We do not find effects on primary school completion and 4-year-college completion, as the coefficients are small and statistically insignificant, though the signs point in the expected direction.

In Figures 6 and 7 we show that the effects on years of education and high school completion are observed only for individuals exposed to hurricanes while in utero. Each figure plots coefficients on indicator variables for hurricane exposure during separate 9-month periods from a single regression. The plots begin with a placebo exposure for the nine months prior to conception and continue up to exposure at age 3. In both cases, we find small and statistically insignificant effects for exposure outside the in-utero period.

6.2 Longevity results

Figures 8 and 9 plot the coefficients for the effects of hurricane exposure on longevity by timing of conception relative to the storm for the D-in-D and mother-fixed-effects regressions, respectively, while Table 5 displays the coefficients in tabular form. In each regression, we estimate separate effects for pre-storm conception (0 to 9 months), in utero exposure, and exposure during the first three 39-week postpartum periods. Our preliminary findings from the mother-fixed-effects regression suggest that exposure to a hurricane during the in utero period, or from birth to age 3, is associated with a 1.5-year reduction in lifespan. This effect represents a substantial 11.5-percent reduction in life expectancy at age 65 (compared to an average of 13 years) for a male born between 1896 and 1900. The effects of postpartum storm exposure are similar in magnitude to the effect of in utero exposure. The coefficient on in-utero exposure is small and insignificant in the D-in-D specification, but exposure to a hurricane from 10 to 18 months of age is associated with a 0.5-year (3.8-percent) reduction in lifespan.

Figure 10 plots the coefficients for hurricane exposure from an alternate mother-fixed-effects specification in which the outcome variable is an indicator for whether an individual died at a younger age than his sibling. We find that hurricane exposure increases the probability of dying at a younger age than ones sibling by 14.7 percent.

7 Robustness checks

7.1 Specification of treatment

Given the nature of hurricanes, we would expect the effects of exposure to storms to dissipate the further an individual was located away from the center of the storm. However, it is not clear how large the exposure area should be since we lack information about the size and intensity of the storm. In our main specification for the effects on educational outcomes, we consider locations within 40 kilometers of the hurricane track to be exposed to the storm and we exclude a "donut hole" region from the estimation sample consisting of counties 40 to 70 kilometers away from the hurricane track. This creates a separation in the degree of exposure to the hurricane between exposed and unexposed zones, the latter consisting of counties 70 to 100 kilometers away from the storm track.

In Table 6 we explore the robustness of the results to different specifications of the hurricane exposure zone and "donut hole" region. First, we restrict the exposure zone to locations within 30 kilometers of the storm track. In column (1) we use the full sample of locations within 100 kilometers of the storm track, while in column (2) we exclude locations in a 30 to 50 kilometer "donut hole" region. Second, we maintain the definition of the exposure zone from the main specification, in which the exposure zone is within 40 kilometers of the hurricane track. In column (3) we also include locations within the 40 to 70km bandwidth "'donut hole" region as part of the unexposed zone, while in column (4) we restrict the "donut hole" to locations 40 to 60 kilometers from the storm track. Third, we extend the exposure zone to include locations up to 50 kilometers from the hurricane track. In column (5) we use the full sample of locations within 100 kilometers of the storm track, while in column (6) we exclude locations in a 50 to 70 kilometer "donut hole" region.

In Panel A, the effects on years of completed schools are generally smaller and insignificant across the robustness specifications, in comparison to the main specification, although we continue to find a significant effect in the case of the 40-kilometer bandwidth treatment definition in the sample with no "donut hole" exclusion. On the other hand, in Panel B, the effects on the probability of high school completion are consistent and similar to our main specification for the 30-kilometer and 40-kilometer bandwidth treatment definition. The magnitudes of the coefficients in the 50-kilometer bandwidth treatment are attenuated due to the likelihood that we include more locations in the exposure zone that were mildly affect or unaffected by the hurricane. These results indicate that the overall effects on years of education are somewhat sensitive to the definition of the treatment, but the effects on the probability of high school graduation are very robust.

8 Conclusion

We have documented large effects of in-utero exposure to hurricanes on the probability of high school completion, in the medium-run, and on longevity, in the long-run. It is not clear, a priori, whether the reduction in lifespan would be associated with an extended or compressed period of morbidity during old age. In either case, the results indicate that there are potentially very large returns over the long run to improving health conditions during fetal development and infancy. Our results suggest a potential policy role for targeted short-run interventions and long-run support to natural disaster victims.

References

Aizer, A., Eli, S., Ferrie, J. and Lleras-Muney, A. (2013). The effects of childhood means-tested cash transfers on mortality. *Working Paper*.

Almond, D. and Currie, J. (2011). Human capital development before age five. *Handbook of Labor Economics*, **Volume 4, Part B**, 1315–1486.

—, Edlund, L. and Palme, M. (2009). Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in Sweden. *The Quarterly Journal of Economics*, **124** (4), 1729–1772.

Antilla-Hughes, J. K. and Hsiang, S. M. (2013). Destruction, disinvestment, and death: Economic and human losses following environmental disaster.

Barker, D. J. P. (1995). The fetal origins of coronary heart disease. *BMJ*, **311**, 171–4.

Boustan, L. P., Kahn, M. E. and Rhode, P. W. (2012). Moving to higher ground: Migration response to natural disasters in the early twentieth century. *American Economic Review*, **102** (3), 238–44.

Crittenden Fuller, S. (2012). The effect of prenatal natural disaster exposure on school outcomes. *Working Paper*.

Currie, J. and Rossin-Slater, M. (2013). Weathering the storm: Hurricanes and birth outcomes. *Journal of Health Economics*, **32** (3), 487–503.

Cutler, D. M., Miller, G. and Norton, D. M. (2007). Evidence on early-life income and late-life health from America's dust bowl era. *Proceedings of the National Academy of Sciences*, **104** (33), 13244–13249.

Deuchert, E. and Felfe, C. (2013). The tempest: Natural disasters, early shocks and children's short- and long-run development. *Working Paper*.

Hill, M. E. and Rosenwaike, I. (2001). The social security administration's death master file: The completeness of death reporting at older ages. *Social Security Bulletin*, **64** (1).

Jayachandran, S. (2009). Air quality and early-life mortality. *Journal of Human Resources*, **44** (4), 916–954.

Lindeboom, M., Portrait, F. and van den Berg, G. J. (2010). Long-run effects on longevity of a nutritional shock early in life: The dutch potato famine of 1846-184. *Journal of Health Economics*, **29** (5), 617–629.

Maccini, S. and Yang, D. (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American Economic Review*, **99** (3), 1006–26.

Simeonova, E. (2011). Out of sight, out of mind? Natural disasters and pregnancy outcomes in the USA. *CESifo Economic Studies*, **57** (3), 403–431.

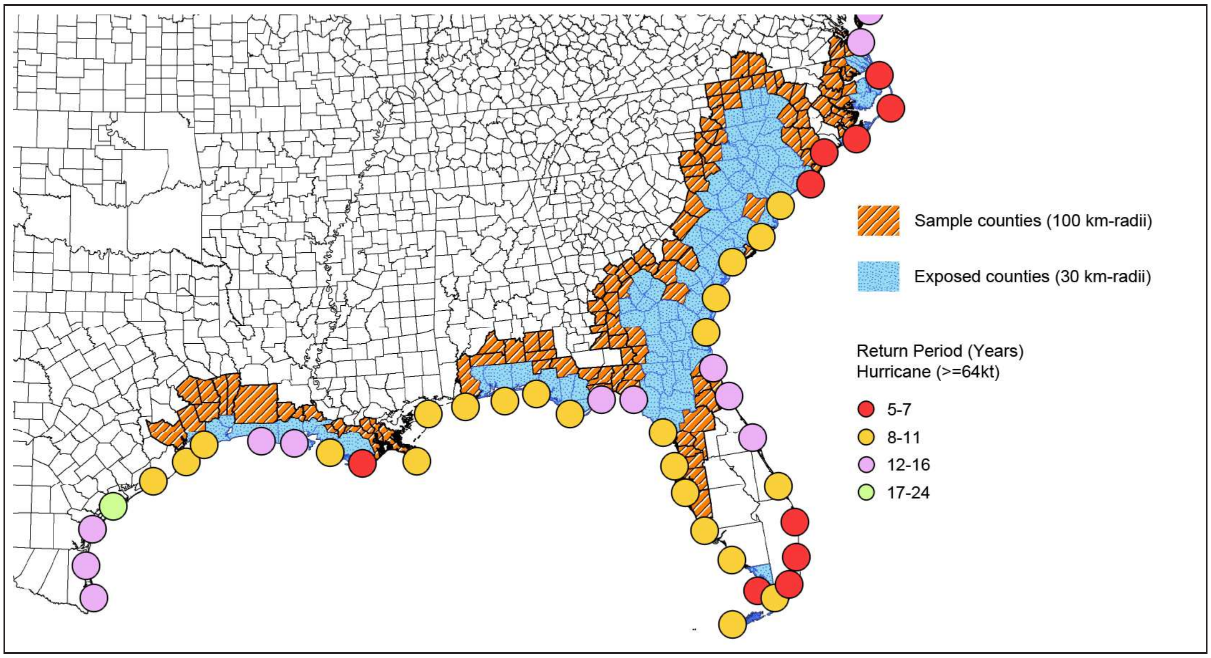
Sotomayor, O. (2012). Fetal and infant origins of diabetes and ill health: Evidence from Puerto Rico's 1928 and 1932 hurricanes. *Economics & Human Biology*, **11** (3).

Torche, F. (2011). The effect of maternal stress on birth outcomes: Exploiting a natural experiment. *Demography*, **48** (4), 1473–1491.

van den Berg, G., Doblhammer-Reiter, G. and 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.

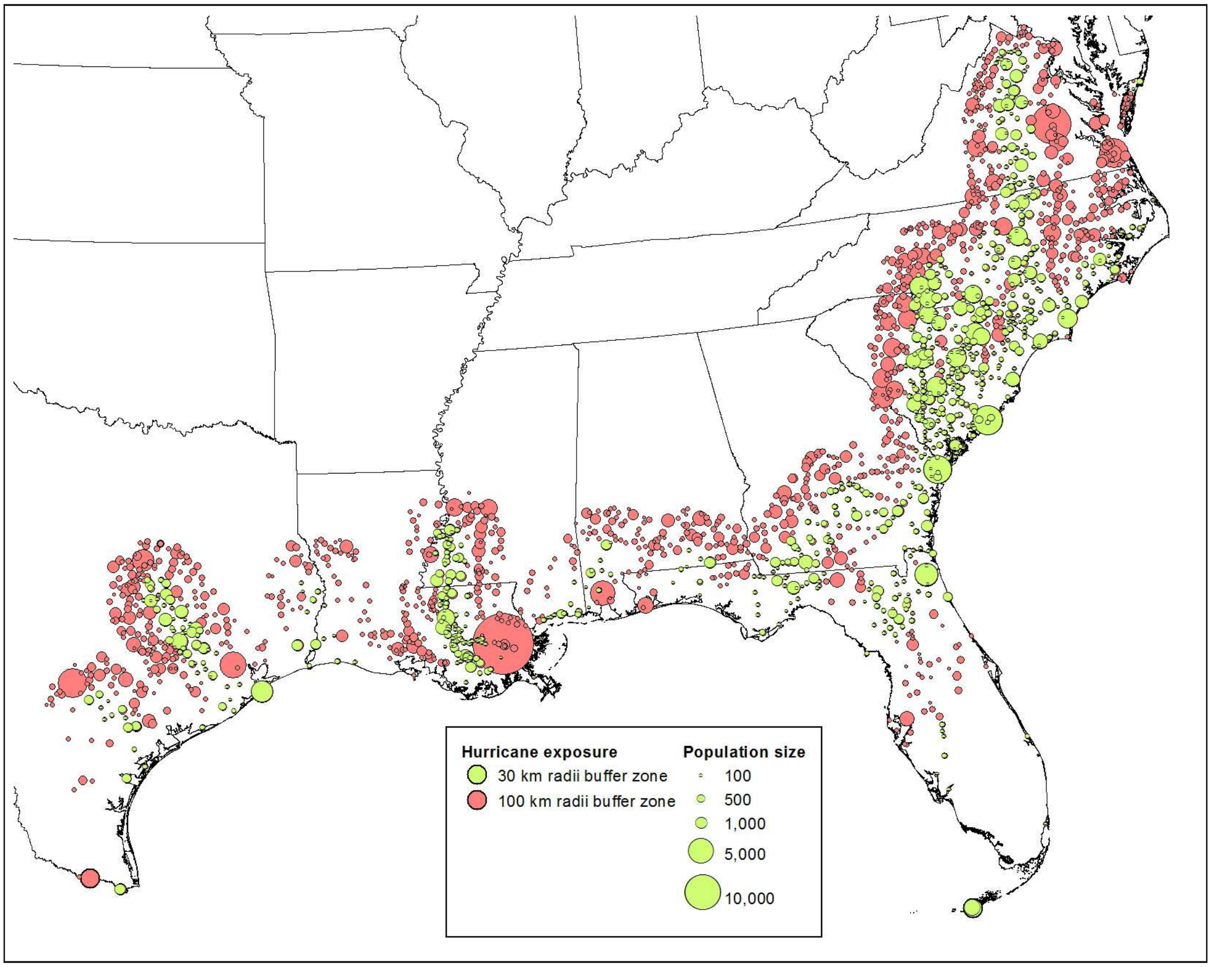
Figures

Figure 1: Counties exposed to hurricanes and all sample counties, 1896-1899



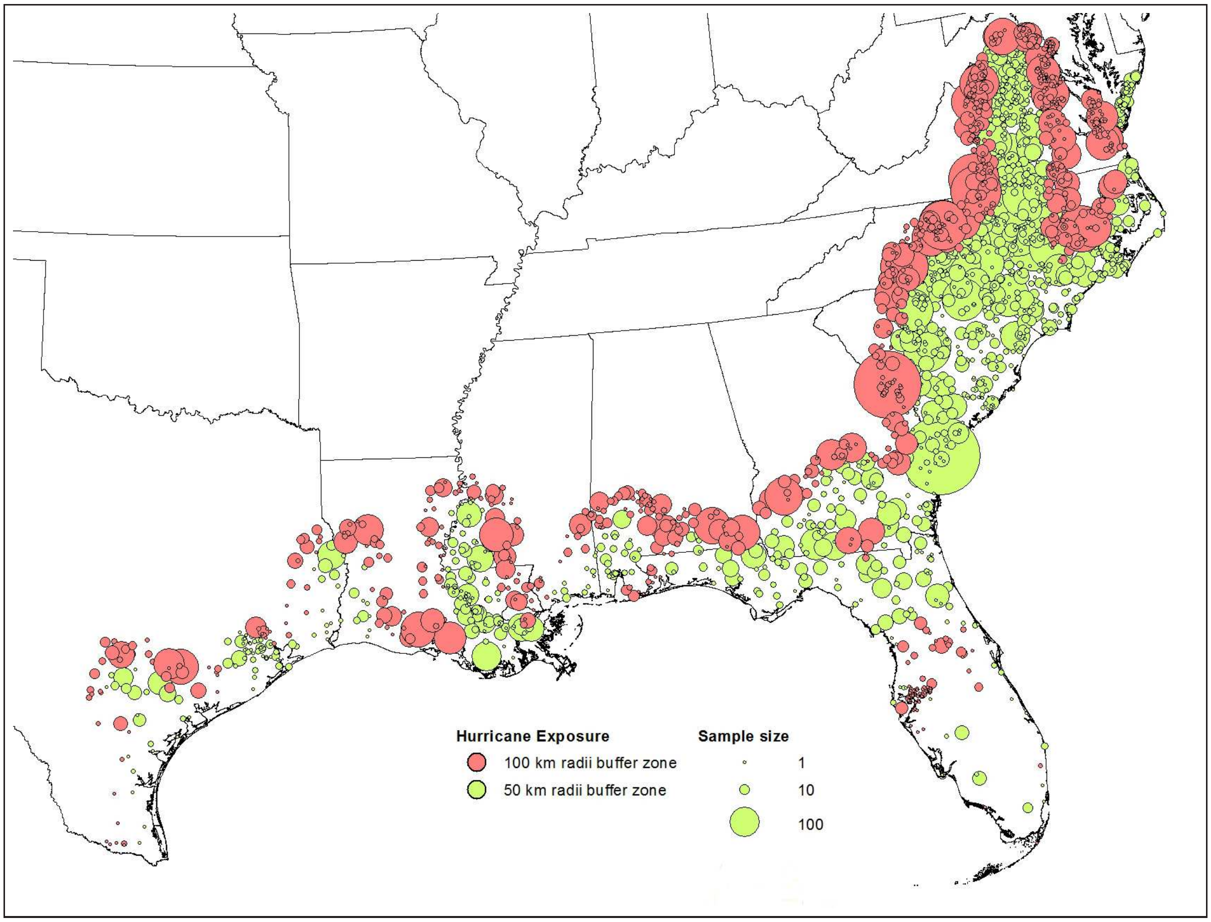
Notes: The counties shaded blue in the figure denote counties that overlap 30-km radii bandwidths around hurricane tracks between 1896 and 1899. We define these counties as being exposed to the hurricane. The counties shaded in orange overlap 100-km radii bandwidths around the storm tracks. All portions of displayed storm tracks are constructed from coordinates at which a minimum of Category 1 strength wind speeds were recorded. The portions of the storm tracks at which a hurricane had dissipated to a tropical storm are not shown.

Figure 2: Population size and hurricane exposure at birth locations



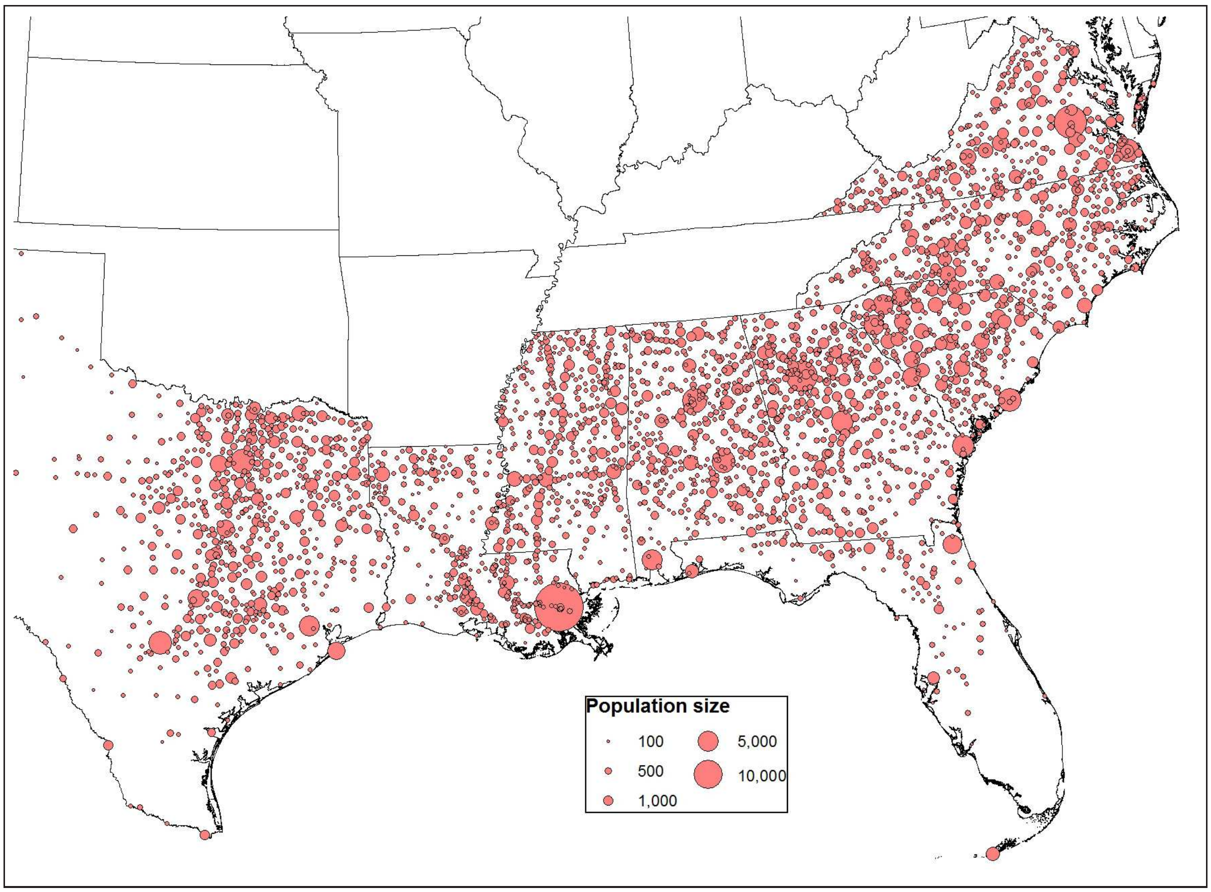
Notes: Figure 2 shows a map of the Southern US. Colored circles on the map denote the coordinates of birth locations in the World War I Draft Registration Cards database. The size of the circles provide a measure of the population at each point. Locations within a 30-kilometer (km) radius of hurricane tracks in the sample are colored green, while locations 30 to 100km away from hurricane tracks are colored red.

Figure 3: Matched sample size and hurricane exposure at birth locations



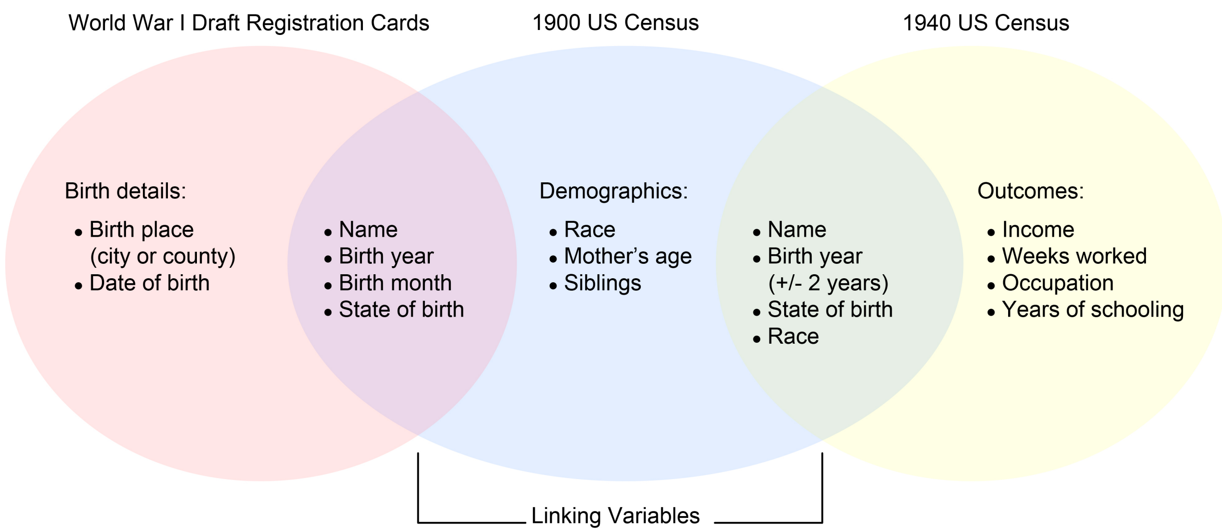
Notes: Figure 3 shows a map of the Southern US. Colored circles on the map denote the coordinates of birth locations in the World War I Draft Registration Cards database. The size of the circles indicates the number of observations matched to the 1900 and 1940 censuses with birth locations at each point. Locations within a 30-kilometer (km) radius of hurricane tracks in the sample are colored green, while locations 30 to 100km away from hurricane tracks are colored red.

Figure 4: Birth locations in WWI Draft Registration Cards



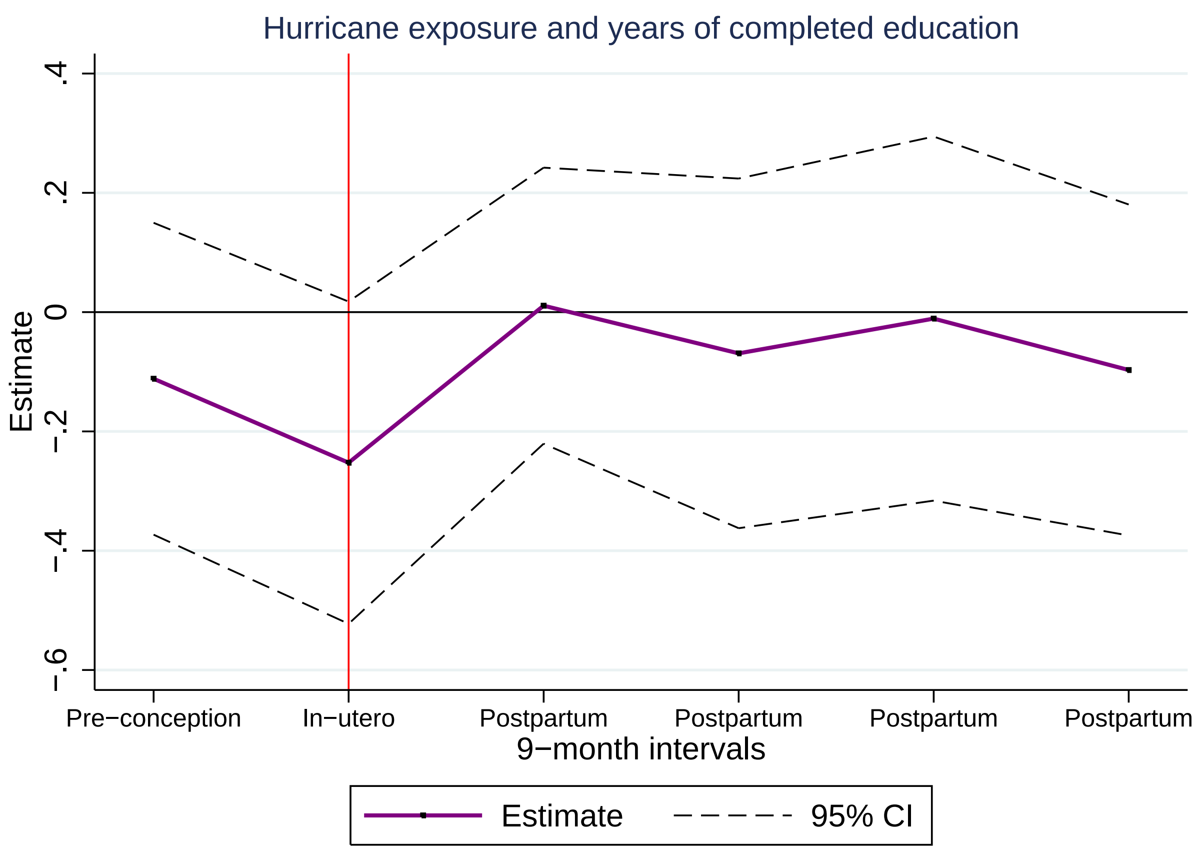
Notes: Figure 4 shows a map of the Southern US. The red circles on the map denote the coordinates of all birth locations in the World War I Draft Registration Cards database. The size of the circles indicates the number of individuals with birth locations at each point.

Figure 5: Data sources and linking variables



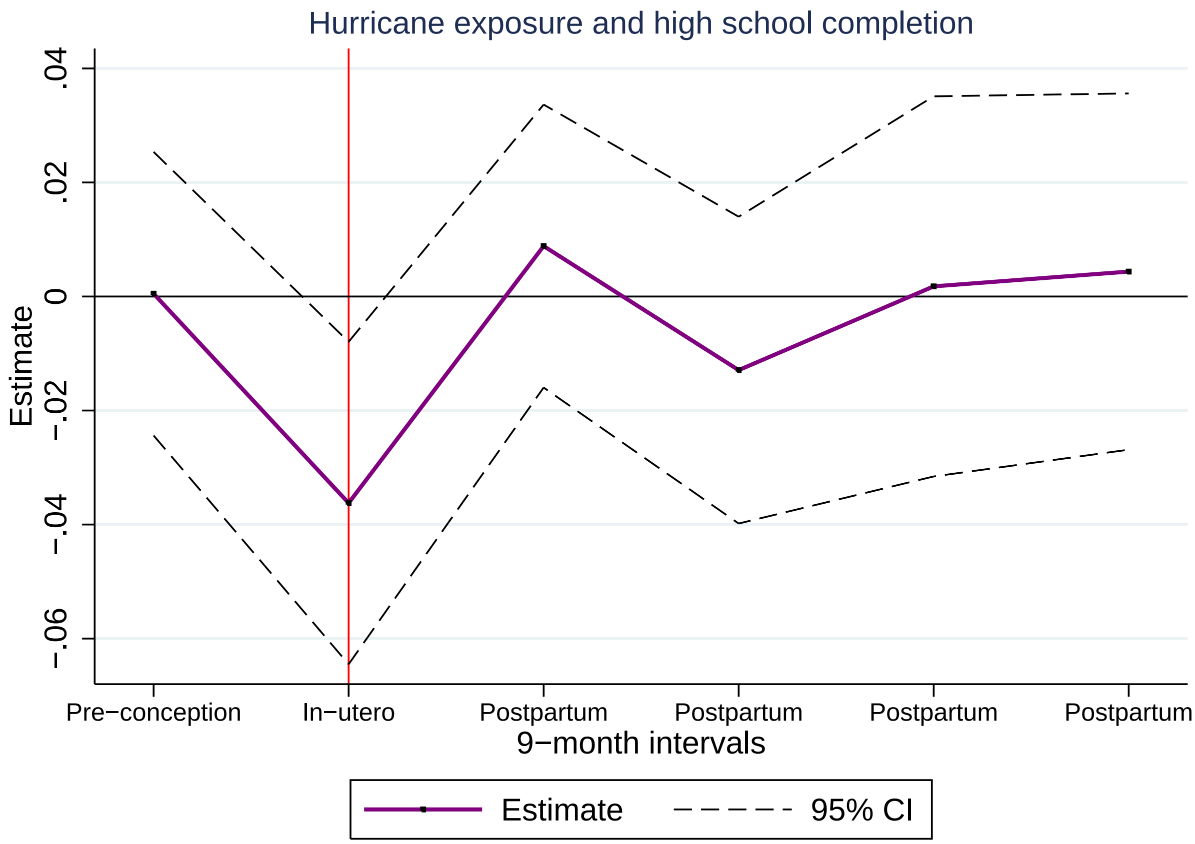
Notes: The figure displays data sources used in the analysis of educational outcomes. From left to right, the World War I Draft Registration Cards are linked to the 1900 U.S. census, and then to the 1940 census.

Figure 6: Coefficients on hurricane exposure: Difference-in-differences model



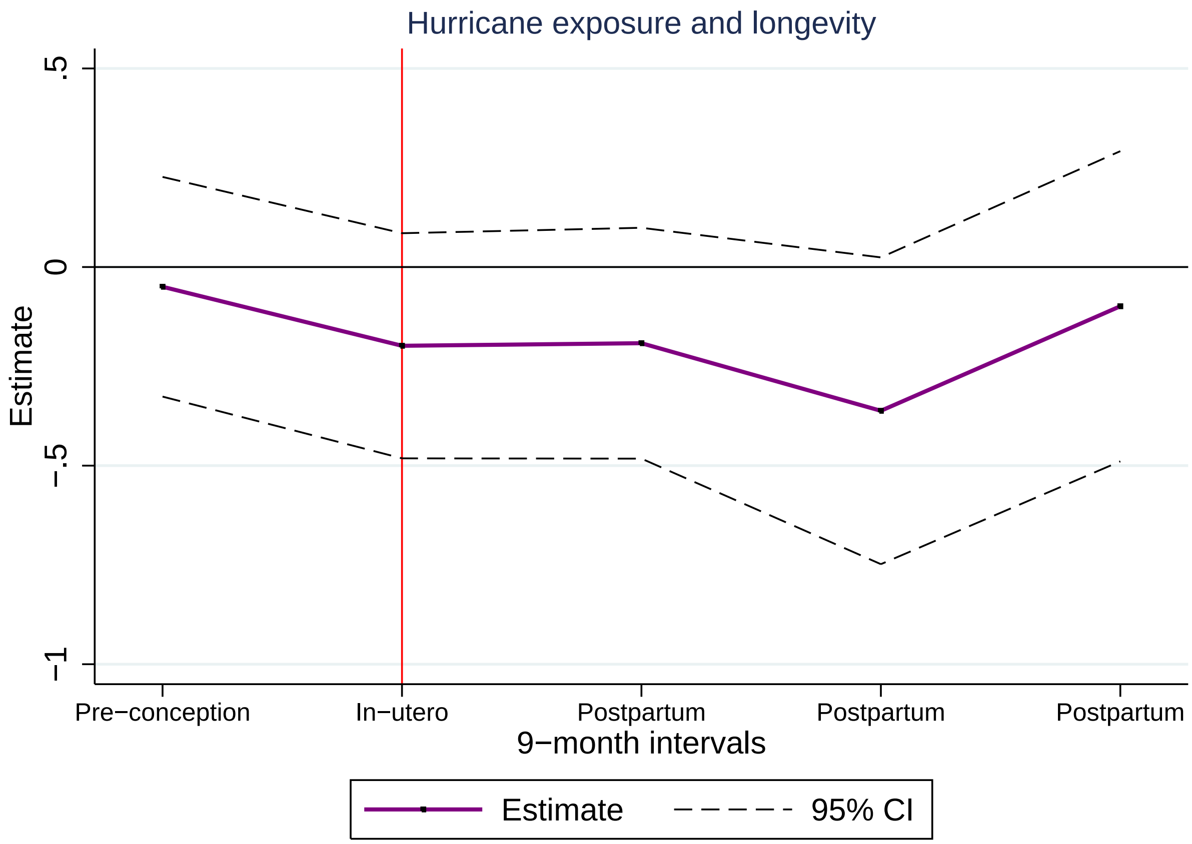
Notes: Coefficients from a single regression estimating Equation 1 with years of completed education as the dependent variable. Standard errors clustered by county of birth.

Figure 7: Coefficients on hurricane exposure: Difference-in-differences model



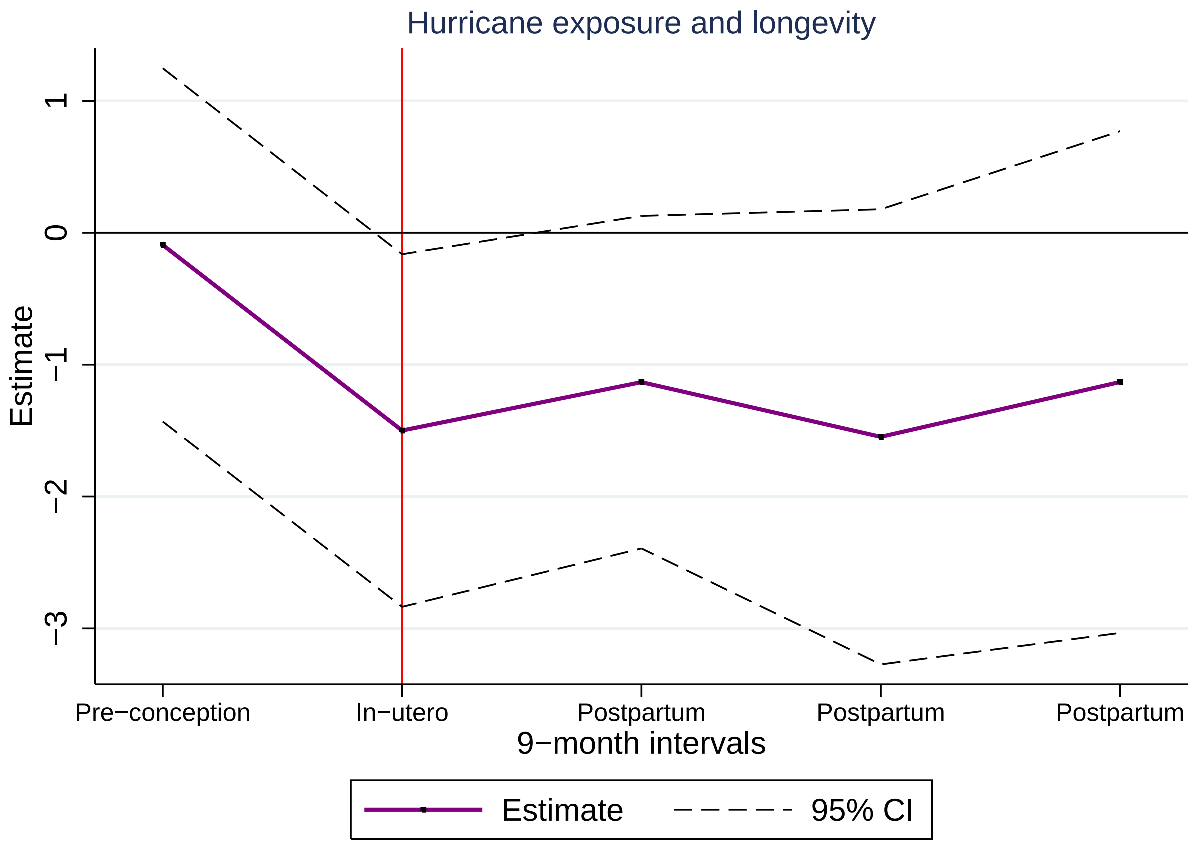
Notes: Coefficients from a single regression estimating Equation 1 with indicator for high school completion as the dependent variable. Standard errors clustered by county of birth.

Figure 8: Coefficients on hurricane exposure: Difference-in-differences model



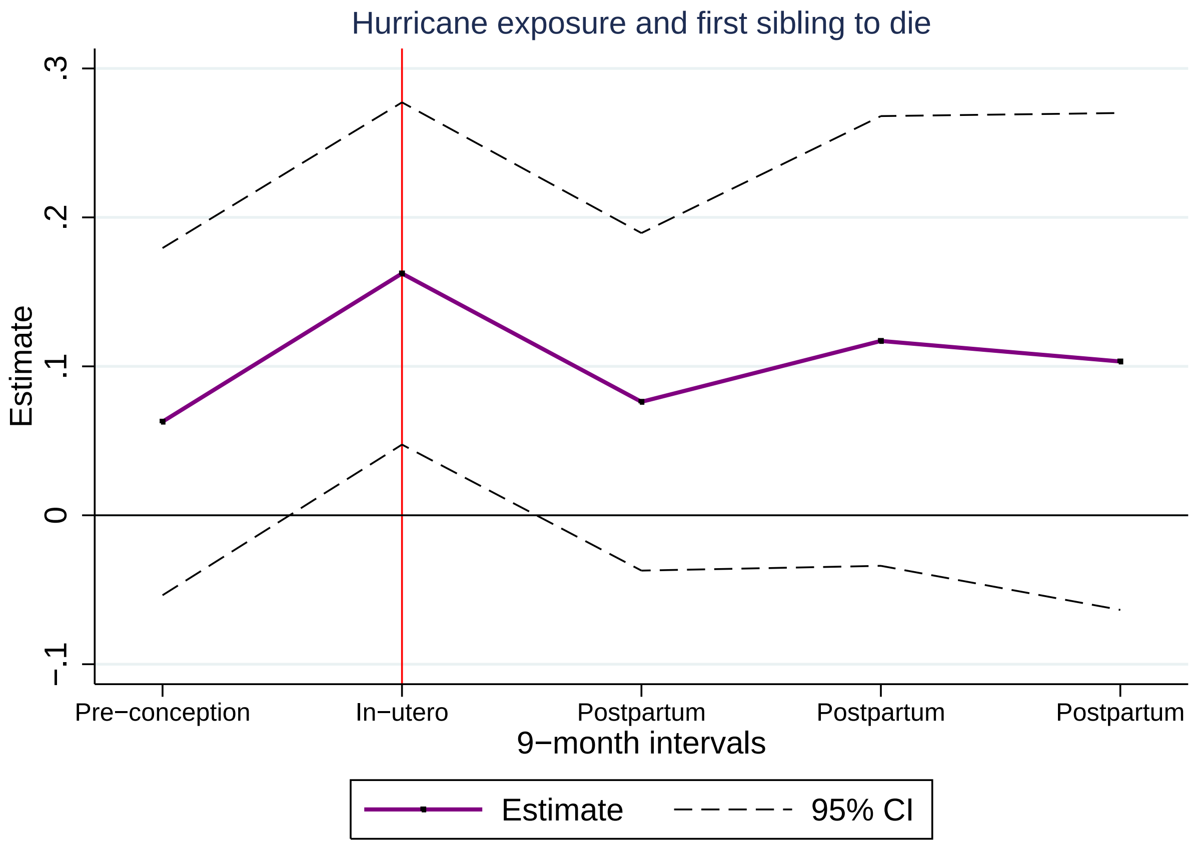
Notes: Coefficients from a single regression presented in Equation 1 with longevity as outcome variable. Standard errors clustered by county of birth.

Figure 9: Coefficients on hurricane exposure: Mother fixed effects model



Notes: Coefficients from a single regression presented in Equation 2 with longevity as outcome variable. Standard errors clustered at mother level.

Figure 10: Robustness check: Mother fixed effects and first sibling to die



Notes: Coefficients from a single regression presented in Equation 2 with dummy variable for first sibling to die as outcome variable. Standard errors clustered at mother level.

Tables

Table 1: Number of hurricanes by category and time period

	Category					All	Major
Time period:	1	2	3	4	5		(3 to 5)
1891-1900	8	5	5	3	0	21	8
1901-1910	10	4	4	0	0	18	4
1911-1920	8	5	4	3	0	20	7
2001-2010	8	4	6	1	0	19	7
1851-2010	113	75	75	18	3	284	96

Source: NOAA

Table 2: Hurricanes with in utero exposure for 1896-1900 birth cohorts

Year	Dates	Category	Average Wind Speed (kph)	Maximum Wind Speed (kph)	Estimated Damage ($)	States Exposed
1896	7/4 – 7/12	2	84.85	160	2,771,777	FL
1896	9/22 – 9/30	3	134.56	205	44,348,432	FL, GA, NC, SC
1897	9/10 – 9/13	1	123.67	140	N/A	FL, LA, TX
1898	8/2 – 8/3	1	77.50	130	N/A	FL
1898	8/30 – 9/1	1	99.09	140	11,087,108	GA, SC
1898	9/25 – 10/6	4	121.56	215	69,294,425	FL, GA
1899	7/28 – 8/2	2	102.73	160	27,717,770	FL
1899	8/3 – 9/4	2	132.77	240	N/A	NC
1899	10/26 – 11/4	2	103.00	175	N/A	NC, SC

Notes: Storms restricted to hurricanes with Category 1 to 5 intensity. Sample includes birth from the [1896w1, 1900w22] cohorts, corresponding to in utero exposure periods ranging between [1895w11, 1899w32], assuming a common 39-week gestation period. Dates of storm and wind speed measures include offshore coordinates. A state is considered exposed to a hurricane if any part of its area overlaps with a 20-km-radii bandwidth around the hurricane track. Estimated damage in 2013 US dollars.

Table 3: Population counts and matching rates by hurricane exposure status

	(1)	(2)	(3)	(4)	(5)
	Southern states	50–100 km radii buffer zone		0–50 km radii buffer zone	
Data sources	Complete count	Not exposed	In-utero exposure	Not exposed	In-utero exposure
	Panel A: Population counts by data source				
1900 U.S. Census	2,516,229				
WWI Draft Registration Cards	1,999,192	93,975	12,447	135,466	11,181
1940 U.S. Census[1]	1,576,452				
	Panel B: Sample sizes in matched datasets [matching rate[2]]				
1900–WWI	635,343	26,356	3,573	33,387	3,106
	[25.2]	[28.0]	[28.7]	[24.8]	[27.8]
1900–1940	749,458				
	[29.8]				
1900–WWI–1940	298,723	12,478	1,684	14,929	1,388
	[11.9]	[13.3]	[13.5]	[11.1]	[12.4]

Notes: The table displays population counts for males born between June 1886 and August 1897 in one of nine Southern states: Alabama, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Texas and Virginia. Column (1) shows the complete population counts for the aforementioned birth cohorts, columns (2) and (4) restrict attention to individuals born in locations within 100 or 50 km, respectively, of a hurricane track that made U.S. continental landfall between 1885 and 1896, and columns (3) and (5) restrict attention to individuals who were in utero while a hurricane passed within 100 km or 50 km, respectively, of their birth location. The hurricane exposure buffer zones are defined based on the detailed place of birth field in the WWI Draft Registration Cards, and thus the exposed population counts are not available for the complete populations enumerated in the censuses.

[1] Given that the 1940 U.S. Census only reported an individual's age in years, we estimate the birth cohorts of interest with individuals aged 43 to 53 on the day of enumeration (April 1, 1940). We allow for up to a 2 year discrepancy in age when matching to the 1900 U.S. Census.

[2] The matching rate in column (1) represent the percentage of the observations in the 1900 U.S. Census that are matched to the other source(s), while the matching rates in columns (2) to (5) are calculated as a percentage of the observations in the WWI records.

Table 4: Main results for effects on educational outcomes

	(1) Years of schooling	Indicators for completion (2) Primary	(3) Secondary	(4) College
In utero exposure	-0.236^{*}	-0.003	-0.038^{***}	-0.008
	(0.129)	(0.018)	(0.013)	(0.009)
Black	-2.804^{***}	-0.338^{***}	-0.155^{***}	-0.034^{***}
	(0.112)	(0.013)	(0.011)	(0.004)
Mother's age under 20	-1.329^{***}	-0.083^{*}	-0.121^{***}	-0.043^{**}
	(0.370)	(0.048)	(0.034)	(0.020)
Mother's age [20,24]	-0.748^{**}	-0.042	-0.071^{**}	-0.019
	(0.371)	(0.046)	(0.032)	(0.020)
Mother's age [25,34]	-0.215	0.007	-0.029	-0.005
	(0.372)	(0.047)	(0.032)	(0.020)
Mother's age [35,44]	-0.114	0.003	-0.017	-0.003
	(0.383)	(0.046)	(0.033)	(0.020)
Mother married	0.370^{***}	0.023	0.027^{**}	0.006
	(0.125)	(0.016)	(0.013)	(0.007)
Death of sibling	-0.417^{***}	-0.029^{***}	-0.040^{***}	-0.011^{**}
	(0.061)	(0.008)	(0.007)	(0.004)
Mean of Y	7.675	0.717	0.193	0.048
Observations	13422	13422	13422	13422

Notes: In column (1) the dependent variable is the years of completed schooling. In columns (2) to (4) the dependent variables are indicators for completion of primary schooling (up to grade 6), secondary schooling (up to grade 12), and college (4 years of post-secondary schooling), respectively. *In utero exposure* is an indicator variable that takes the value of one if a hurricane track passed within 40 kilometers of the individual's county of birth during the in-utero period. *Black* is an indicator for an individual whose race is black, with individuals reporting a race of white as the excluded category. Age dummies reflect the mother's age when giving birth to the individual with the excluded category being mothers aged 45 to 49. *Death of sibling* is an indicator that equals one for a mother that reports a greater number of children ever born than children alive at enumeration in the 1900 census. In addition to the variables on display, the regressions control for birth order, county of birth, conception year, and conception month fixed effects. Standard errors are clustered by county of birth.

Table 5: Main results for effects on longevity

	Longevity		Shorter life
	(1) D-in-D	(2) Sibling FE	(3) Sibling FE
Pre-conception (0 to 9 months)	−0.050	−0.092	0.070
	(0.141)	(0.683)	(0.059)
In utero (9 months)	−0.198	-1.500^{**}	0.158^{***}
	(0.145)	(0.682)	(0.059)
Post-birth (0 to 9 months)	−0.192	-1.133^{*}	0.077
	(0.148)	(0.643)	(0.058)
Post-birth (10 to 18 months)	-0.362^{*}	-1.547^{*}	0.118
	(0.197)	(0.880)	(0.077)
Post-birth (19 to 27 months)	−0.099	−1.132	0.116
	(0.199)	(0.970)	(0.085)
Mean of Y	79	79	1
Observations	59270	5172	5172

Column (1) displays coefficients on the hurricane exposure indicators from estimating Equation 1, while Columns (2) and (3) show coefficients from estimating Equation 2 respectively. The coefficients are plotted in Figures 8 to 10. In Columns (1) and (2) the dependent variable is the number of years lived after the age of 65, while in Column (3) it is an indicator for the first to die in a sibling pair. Standard errors are clustered by county in Column (1) and by mother in Columns (2) and (3).

Table 6: Robustness to specification of donut hole region

	Exposure 0-30km		Exposure 0-40km		Exposure 0-50km	
	(1)	(2)	(3)	(4)	(5)	(6)
	No donut	Donut 30-50km	No donut	Donut 40-60km	No donut	Donut 50-70km
			Panel A: Effects on years of schooling			
In utero exposure	−0.160	−0.106	−0.218*	−0.187	−0.106	−0.142
	(0.151)	(0.147)	(0.123)	(0.125)	(0.119)	(0.123)
Mean of Y	7.667	7.667	7.667	7.699	7.667	7.663
			Panel B: Effects on probability of high school completion			
In utero exposure	−0.038**	−0.035**	−0.036***	−0.033**	−0.022	−0.027*
	(0.017)	(0.017)	(0.013)	(0.013)	(0.013)	(0.014)
Mean of Y	0.189	0.189	0.189	0.193	0.189	0.192
Observations	19691	15884	19691	15371	19691	15602

Notes: In each column of Panel A the dependent variable is the years of completed schooling, while in the columns of Panel B it is an indicator for completion of high school (grade 12). In columns (1) and (2) the treatment variable is an indicator for in-utero exposure to a hurricane, where the exposure zone is defined as locations within 30 kilometers of the storm track. In columns (2) and (3) the exposure zone is extended to a 40-kilometer bandwidth, while in columns (5) and (6) it is extended further to a 50-kilometer bandwith. Columns (2), (4), and (6) exclude locations from the sample that fall within the distance band specified in the the column header. See Table 4 for a description of the other control variables included in the regressions.