Online Appendices for:

"Increased mortality of white Americans and a decline in the health of cohorts born after World War II"

by Nicholas Reynolds

July 31, 2024

Appendix A: Lack of geographic variation

I show that the cohort-specific trend break in log mortality rates documented above is remarkably widespread across the United States, suggesting that the associated health decline in similarly widespread. To do so, I estimate the shared cohort-specific trend-break model of Section 4 separately for different states and regions of the United States.

First, I examine the location and size of the trend break by Census region. For each of the four regions, I estimate the trend break model based on equation 2 with a full set of year fixed-effects and a separate linear age effect for each year (similar national results are in column 2 of Table 1). I again follow Hansen (2000) and the procedure described above.

Panel A of Appendix Table 3 shows the results for white women. The precise cohort at which the trend break is estimated to have occurred varies only slightly across the four Census regions — from 1946 in the West to 1950 in the Midwest, with the estimates in the South and Northeast in between at 1948 and 1949 respectively. The average size of the estimated cohort break — the average value of $\delta_{2,c}$ across all years — ranges from a low of .018 in the West to a high of .024 in the Northeast. For all regions, the bootstrap procedure to test the null hypothesis that no trend break occurs suggests a P-value of less than .001.

Panel B shows analogous results for white men. The cohort at which the trend break is estimated to have occurred again varies only slightly across the four Census regions — falling at 1942 in the West, 1946 in the Midwest and South, and 1944 in the Northeast. The average size of the estimated cohort break for men are *remarkably* similar across the four regions. This size is estimated to be identical up to 3 digits — at .026 — for the Northeast, South, and West. While the estimate for the Midwest is not far off at .029. For all regions, the bootstrap procedure to test the null hypothesis that no trend break occurs suggests a P-value of less than .001

To further explore potential geographic variation in the cohort-specific trend break I next examine it separately for each of the 50 states in the U.S. Given the smaller sample size at the state-level, I impose the *location* of the cohort-specific trend break in each state to match that at the national level. That is, I estimate separately for each state the trend break model based on equation 2 but fix

γ_c to be 1949 for women and 1946 for men. I again use a specification with a full set of year fixed-effects and a separate linear age effect for each year, and estimate by weighted least squares — using the implied variance of log mortality as weights. For each state I then calculate the average size of the estimated cohort break — the average value of $\delta_{2,c}$.

Appendix Figure 2 shows maps and histograms of the distribution of these estimated break sizes for the 50 states — and demonstrates a surprising lack of variation in the size of the estimated breaks across states. No obvious regional patterns are apparent in the maps for either sex — the trend break is widespread across the United States. Further, *all 50 states* have estimated trend breaks which are positive in magnitude and greater than .01 for women and men. Estimates for women in all states are between .005 and .045, and 30 out of 50 states have estimated break sizes between .015 and .025. For men estimates range from .01 to .055, and 32 out of 50 states have break sizes between .025 and .035.

Appendix Figure 3 shows a scatter plot between the break sizes of men and women, and reveals a positive association. States with large breaks for men tend to also have large breaks for women. Alaska and Vermont stand out as states with large breaks for both men and women. On the other extreme, California and Florida have particularly small estimated breaks for both sexes. This positive association suggests that a single factor varying at the state-level may be driving health differences for both men and women. A regression of mens break sizes on womens break sizes confirms the positive relationship shown in the scatter plot. Using the estimated variance of the female break sizes from the first-step as weights I perform a second-step, state-level regression. The estimated coefficient on male break size is .659, with a standard error of .079 and a corresponding t-statistic of 8.29.

Appendix B: Early-life mortality: Additional details and results

This appendix provides more details and additional results based on the mortality rates of White Americans under age 30.

As described in the data section, I combine death counts by age-sex-race from historical vital statistics volumes from 1933 to 1958 and counts from multiple cause of death data from 1959 with population estimates from the Census Bureau and SEER.[1] The historical data reports mortality by exact year of age up to age 4, and then only reports mortality in 5-year age bins for older ages. For consistency I aggregate post-1959 data into the same age bins.

I again use the framework of Hansen (2000) to test for the existence of and estimate the location

[1] I use Census Bureau population estimates for 1933-1968 and SEER population estimates from 1969 on.

of trend breaks by cohort. I estimate models of the following form, separately for particular ages or age groups a:

$$ln(mort_{apc}) = \beta_c^a c + \delta^a \cdot (\gamma^a - c) \cdot 1_{c \geq \gamma^a} + \mu^a + \epsilon_{apc} \quad (1)$$

where a denotes age, p denotes period (eg. year), c denote cohort; and $ln(mort_{apc})$ denotes the log-mortality rate of individuals age a, in period p, and from cohort c.[2] The parameters β_c^a represents a linear trend in cohort for log mortality at age a. I then allow a trend break by cohort — thereby letting the affect of cohort have a piecewise linear form. The size of the trend break is represented by δ^a. The precise cohort at which the trend break occurs is treated as unknown and a parameter to be estimated, γ^a.

I calculate the cumulative mortality rate by first assuming that mortality is uniformly distributed by age within the five-year age bins and then decrementing each cohorts mortality using the single age mortality rates. This will introduce some measurement error in mortality for ages over 5, potentially smoothing and understating the size of the break because cohorts will inherit some of the mortality of nearby cohorts in the same age bin. For example, when the 1946 cohort is 9 it will share the 5-9 age bin with the 1947-1950 cohorts and will be assigned an average mortality which includes these cohorts potentially elevated mortality.

Appendix Figure 4 disaggregates these results by age, and shows broadly that this cohort specific trend break of mortality is evident across ages from infancy through childhood. It reports the results of estimating cohort trend break models separately for log mortality rates by single years of ages below age 4, and for 5-year age bins up to age 19. For all ages between 0 and 19 a trend break is estimated to occur somewhere between the 1940 and 1951 cohort, and the trend breaks are all estimated to have positive sign: implying a slowing of improvements in mortality after that cohort. The confidence intervals for the trend breaks for the single ages 1 to 4, as well as those for 5-9, for females all include 1949 and those for men all include 1946 — consistent with the estimates in Figure 7 and for adult mortality earlier in the paper. Those for infant mortality and for ages 10-14 and 15-19 differ slightly: with infant mortality for both sexes estimated to break a few cohorts later, and that for preteens and teens estimate to occur a few cohorts earlier (but with large confidence intervals).

The patterns by birth cohort of log mortality of white Americans in their twenties are different than that described above for earlier ages. However, the mortality rate at these ages is mostly be driven by external deaths (see eg. Figure A1 in Schwandt and Von Wachter (2020)) which seem

[2] For the five-year age bins the cohort is defined based on the youngest age in the bin. Eg. for the 5 to 9 year old age group, cohort c is defined as $p - 5 - 1$. Therefore, if the mortality rate has a trend break in the year when some cohort c' first enters a bin, the break will be estimated to occur at cohort c'.

likely to be less informative about the underlying "health" of these cohorts. Mortality rates at ages 20 to 29 do not exhibit any evidence of a health decline for the post-late-1940s cohorts, like that documented at older ages in earlier sections and potentially suggested at younger ages in this section. There are some fluctuations in mortality for these cohorts but they appear to be driven by period-specific phenomenon — for example there is a sharp increase in mortality from homicide, suicide, and drug poisonings beginning in the late 1960s, and mortality from HIV/AIDS sharply increasing after 1980.[3]

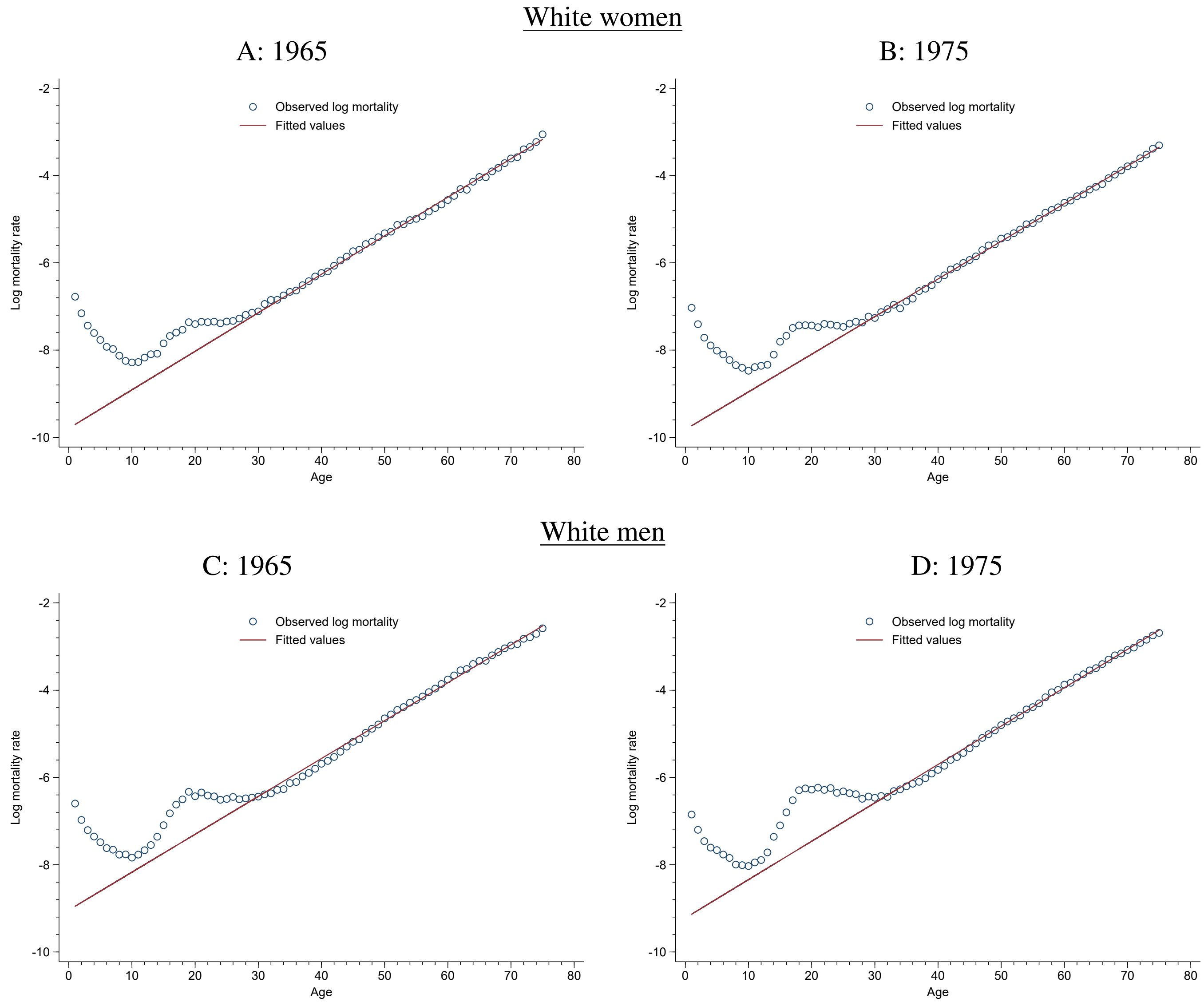
Appendix C: Additional details on period life expectancy calculation

I estimate Equation 2 for log mortality rates in 2019, using the sample and methodology underlying Figure 5. I then form adjusted log mortality rates which remove the relative cohort health decline by subtracting $\hat{\hat{\delta}} \cdot (\hat{\gamma} - c) \cdot 1_{c \geq \hat{\gamma}}$ from the observed log mortality rates in the sample. I then exponentiate these adjusted mortality rates and take the ratio with observed mortality, to form a mortality adjustment ratio r_a

I then use CDC life tables for non-Hispanic men and women in 2019 from Arias and Xu (2022) as a baseline. I form adjusted q_x values by multiplying q_x for each age in the above estimation sample by the mortality ratio r_a described above. I then use standard life table methods to calculate life expectancy using the adjusted q_x values.

[3]See Shahpar and Li (1999) for a discussion of the period-specific increases in the 1960s and 1970s in homicide mortality. My own informal, descriptive analysis suggests that there were sharp, period-specific (ie. occurring at the same year for all single-ages of the twenties) increases in the 1960s in homicide, suicide, and drug poisonings which were large enough to drive all-cause mortality increases. Results available by request.

Appendix Figure 1: Examples of fit of Gompertz curve for select pre-sample years

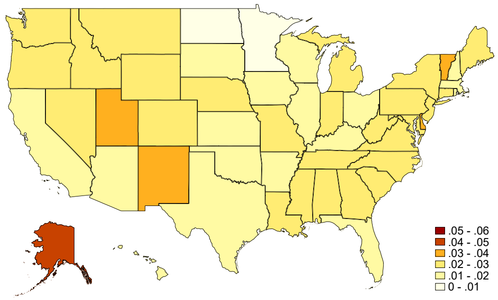


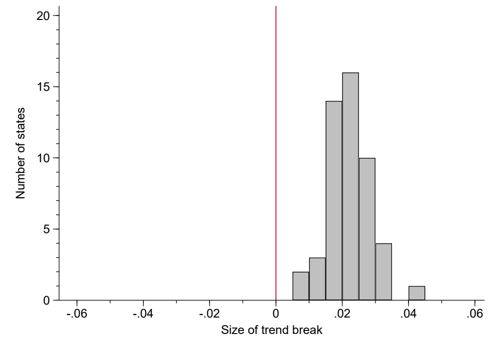
Each plot shows the log mortality rate of white men by age for the year listed, for 1930 to 1965 cohorts. Red circles show the observed log mortality rate by single year of age. The solid blue line shows plots the piecewise-linear, trend-break model estimated by weighted-least squares based on equation 2. The vertical gray line shows the age/cohort of the estimated break in trend. The dotted blue line extrapolates the linear trend for cohorts born before the break to post-break cohorts.

Appendix Figure 2: Little variation across states in size of cohort-specific trend break in log mortality

A: White women

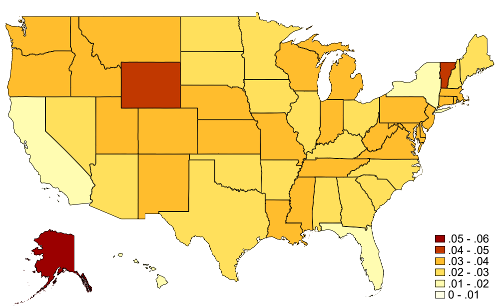
break at 1949 cohort

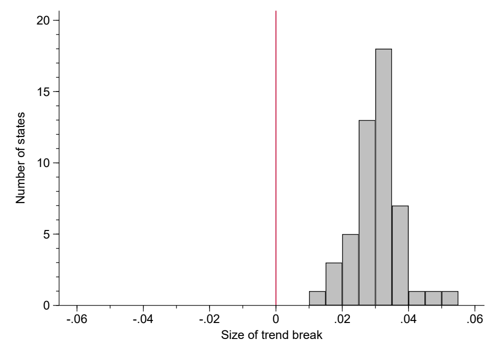




B: White men

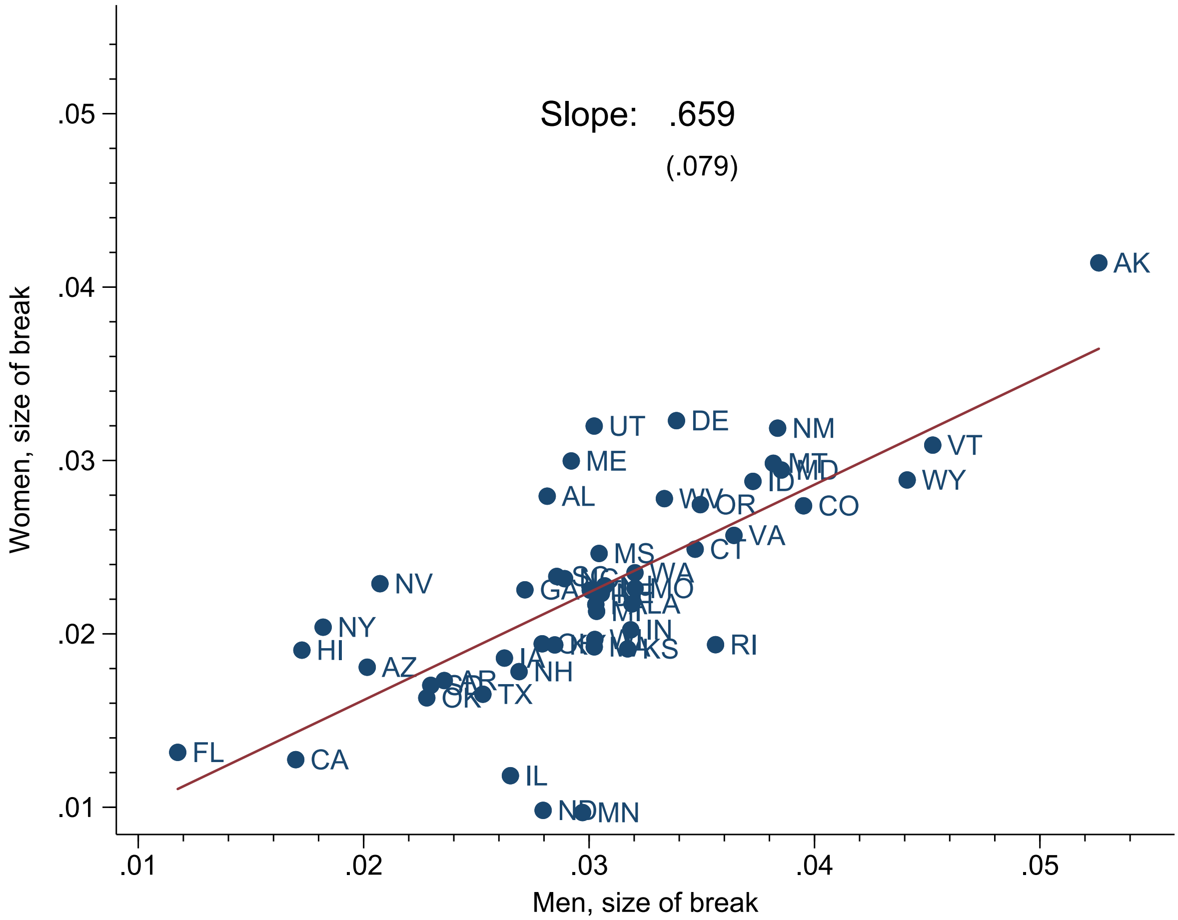
break at 1946 cohort





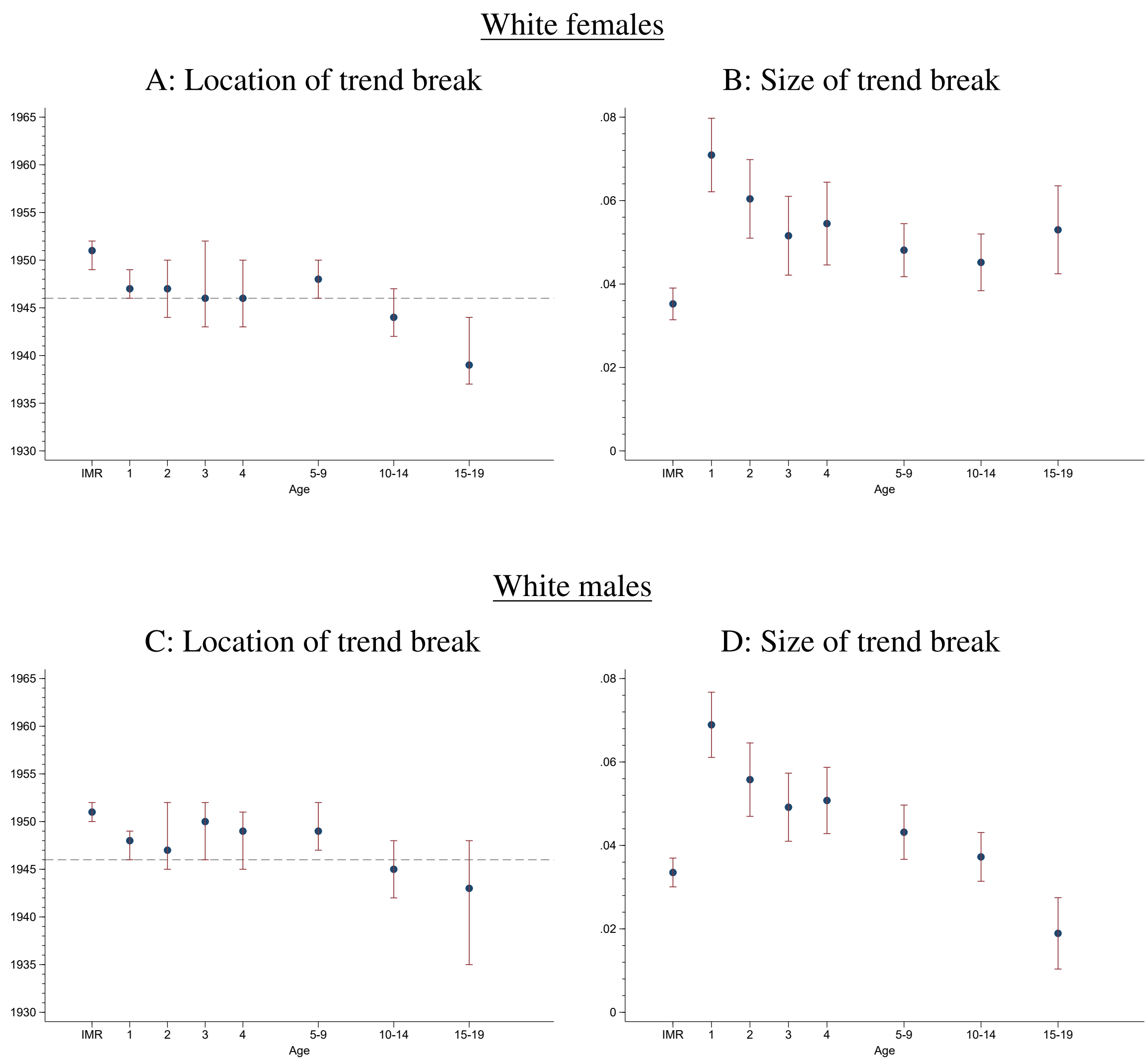
This figure shows the variation across states in the size of cohort-specific trend breaks in the log mortality of white women and men. All figures are based on separate estimation by state of cohort-specific trend break models of the log mortality of white women or men. Each model is based on equation 4 including a full set of year fixed-effects and a separate linear age effect for each year. The location of the trend break γ is treated as known — 1946 for men and 1949 for women — and estimation is done by weighted least squares. The sample includes the years 1985-2015, ages 30-75, and cohorts born from 1930-1970. For each state I calculate the average value of the trend break $\delta_{2,c}$ across all years. The maps display the values of these average trend break sizes for each state. The histograms show the distribution of these average trend break sizes across the 50 states.

Appendix Figure 3: Relationship between state-level size of cohort-specific trend break in white log mortality for women and men



This figure shows the relationship across the 50 states between the size of cohort-specific trend breaks in log mortality for white women and those for white men. The first step is separate estimation by state of cohort-specific trend break models of the log mortality of white women or men. Each model is based on equation 4 including a full set of year fixed-effects and a separate linear age effect for each year. The location of the trend break γ is treated as known — 1946 for men and 1949 for women — and estimation is done by weighted least squares. The sample includes the years 1985-2014, ages 30-75, and cohorts born from 1930-1970. For each state I calculate the average value of the trend break $\delta_{2,c}$ across all years. The above figure plots this average value for each state for men versus the average value for women in the same state. The second step is a regression with the estimated break sizes of women as the dependent variable and that of men as the independent variable. The variance of women's estimated break size from the first step are used as weights in the second step.

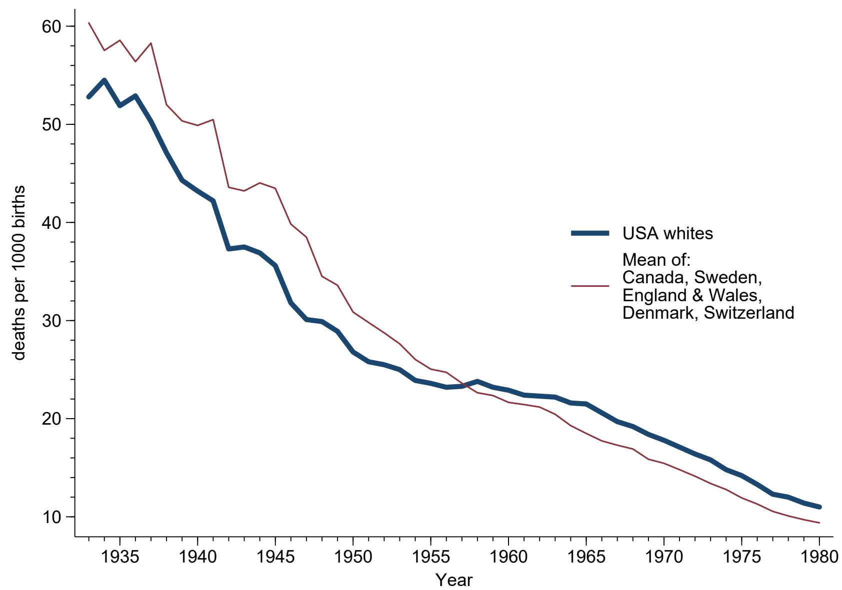
Appendix Figure 4: White childhood log mortality trend break estimates by age



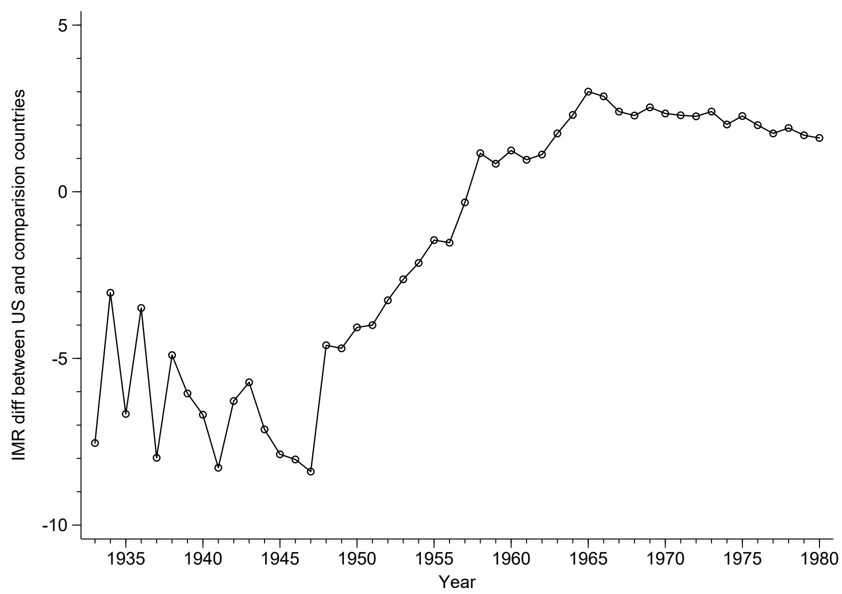
These figures show the results of estimation of the trend break model in equation 5, with the log mortality rate of white females or males by cohort and age as the dependent variable. A separate model is estimated for each age/age-group shown by least squares, following the approach outlined in Hansen (2000). IMR refers to the log of the infant mortality rate, ages 1 to 4 refer to the log mortality rate for single ages between 1 and 1, while the remaining points refer to the log mortality rate for the listed age bin based on decrementing the crude death rates for single years of age, and then taking the natural log (more detail in text). For the five-age bins cohort is defined based on the youngest age in the bin. Panels A and C report estimates of the cohort at which the break is located, $ga\hat{m}ma^a$ 99 percent confidence intervals. Panels B and D report estimates and 95 percent confidence intervals of the size of the break, δ^a of the The underlying rates are calculated based on birth and death counts from Vital Statistics volumes 1933-1959, the Multiple Cause of Death File 1959-2000, and population estimates from SEER and Census.

Appendix Figure 5: Infant mortality rate of White Americans in comparison to other English-speaking and European countries

A: IMR in US and comparison countries

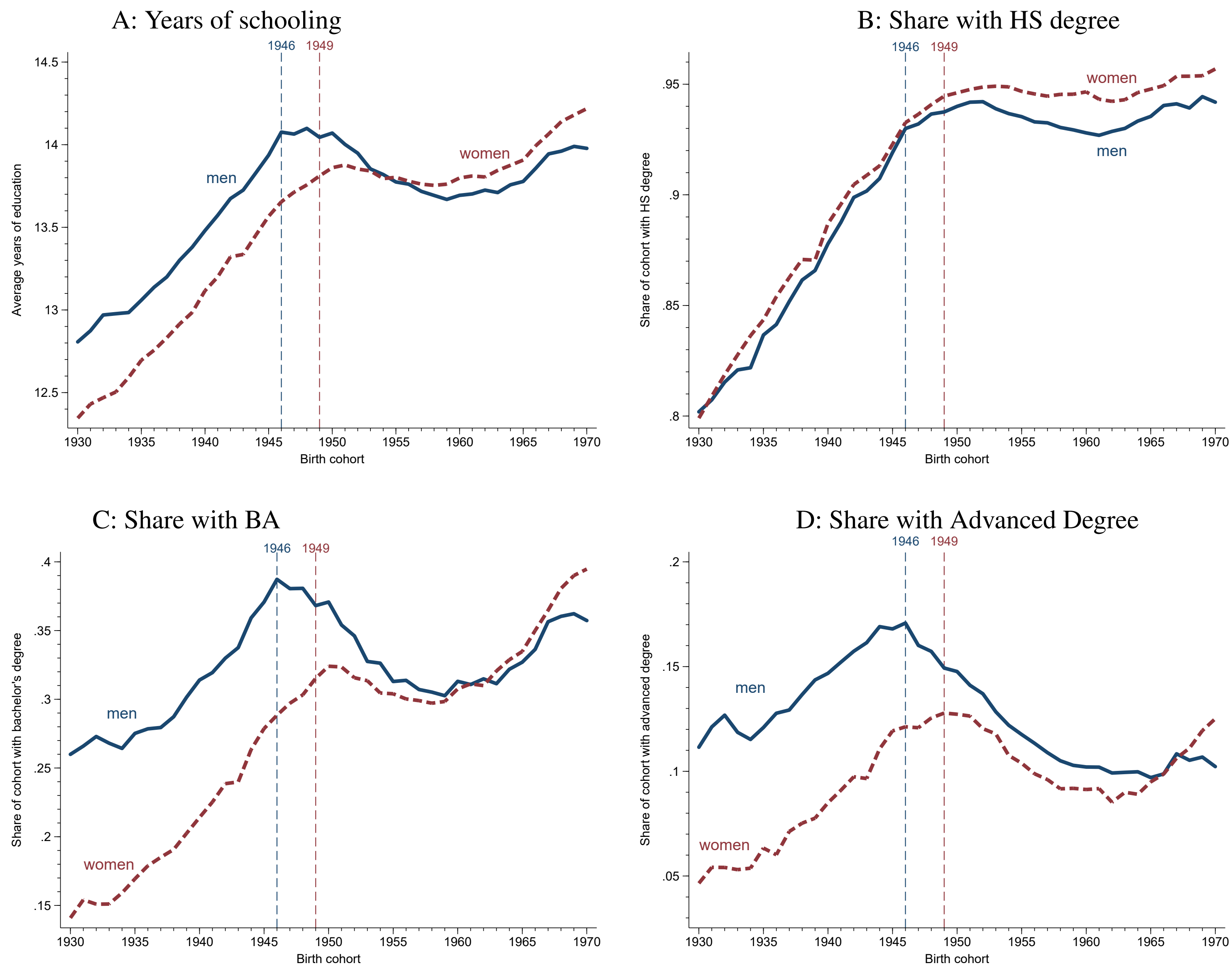


A: Difference in IMR between US and comparison countries



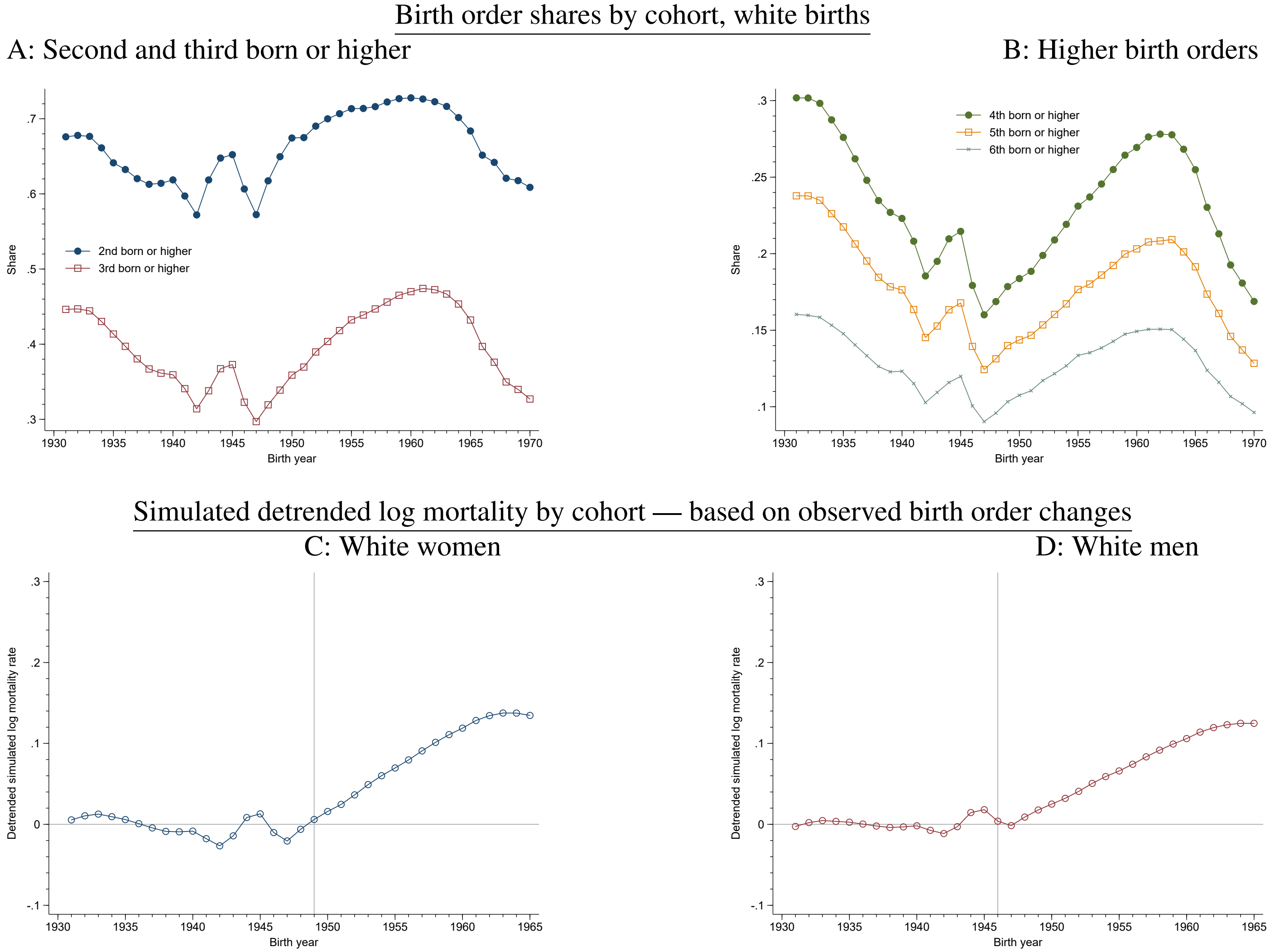
Panel A plots the infant mortality rate of white Americans based on vital statistics volumes and microdata, and the mean infant mortality rate across Canada, Sweden, England and Wales, Denmark, and Switzerland from the Human Mortality Database. Panel B plots the difference between these two series.

Appendix Figure 6: Cohort decline in educational attainment for white men and women



Data is from CPS Merged Outgoing Rotation Group and includes white men and women age 25-75 in years 1990-2018. Panel A plots the average years of schooling by birth cohort — approximated based on 16 educational categories. Panels B-D plot respectively the share of each birth cohort with a high school or GED degree, a bachelor's degree, and an advanced degree.

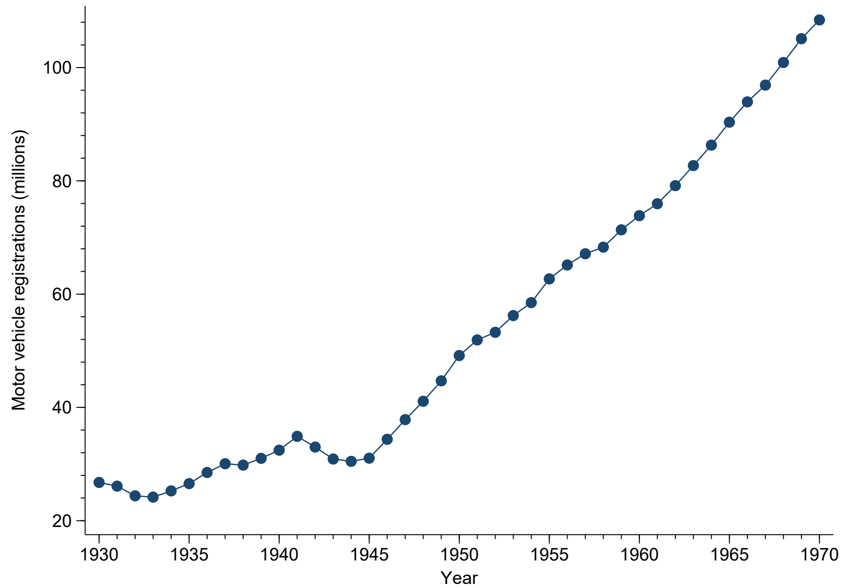
Appendix Figure 7: Birth order changes across white cohorts and simulated effect on log mortality



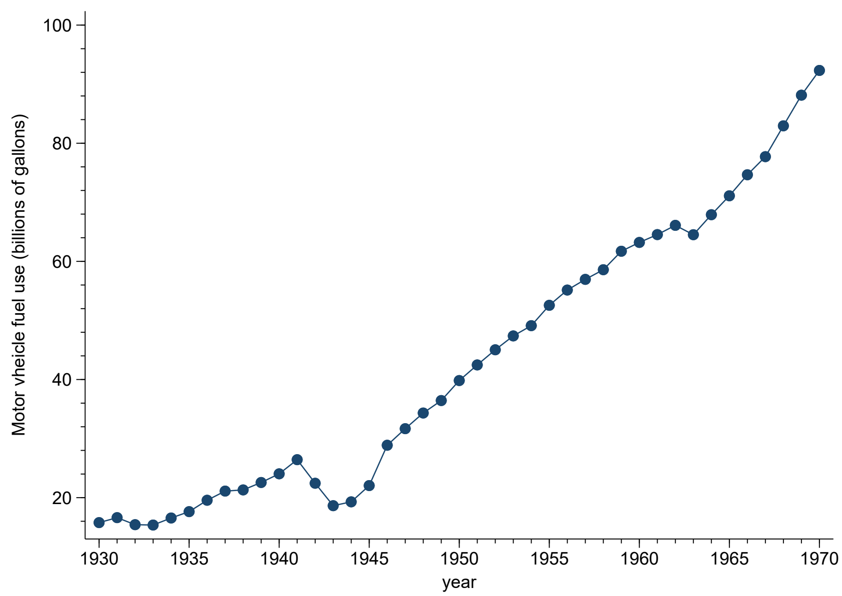
Panels A and B show the share of white births by cohort which are of the listed birth order (parity) or higher. Data for 1930-1939 are digitized from Vital Statistics reports, and for 1940-1970 are calculated from "U.S. Cohort and Period Fertility Tables, 1917-1980" compiled by Robert D. Hauser and available from the Office of Population Research at Princeton. Panels C and D report detrended simulated log mortality rates to show the impact of these birth order trends on log mortality rates by cohort. The simulation uses observed birth order shares, odds ratio estimates of the impact of birth order on mortality from Barclay and Kolk (2015), and observed mortality rates at age 40 of the 1949 and 1946 cohorts of white women and men, respectively. More details in text.

Appendix Figure 8: Motor vehicle registrations and fuel use

A: motor vehicle registrations



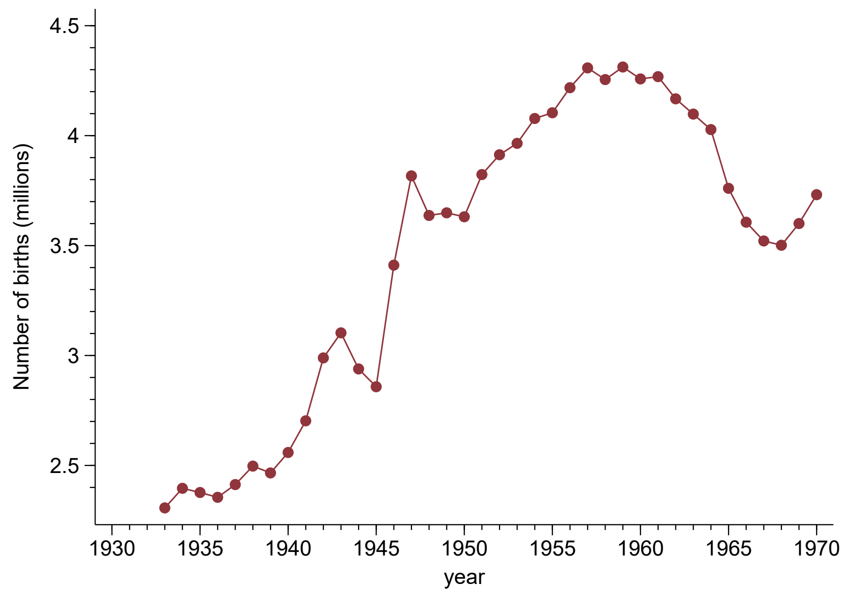
B: Fuel usage by motor vehicles



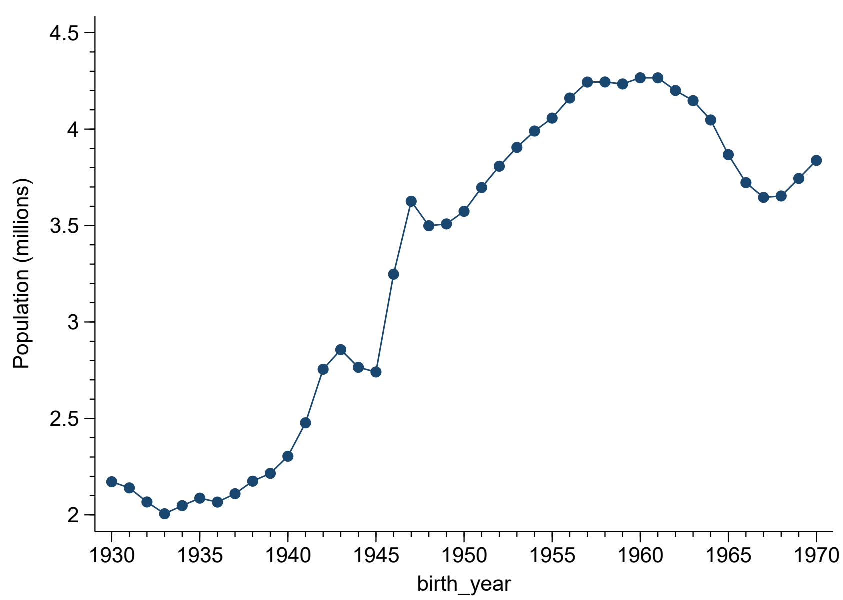
This figure shows annual time series of a) motor vehicle registrations and b) fuel usage by motor vehicles, for 1930 to 1970 in the United States. Both series come from Historical Statistics of the United States US Census Bureau (1975).

Appendix Figure 9: Cohort size

A: Number of births by year

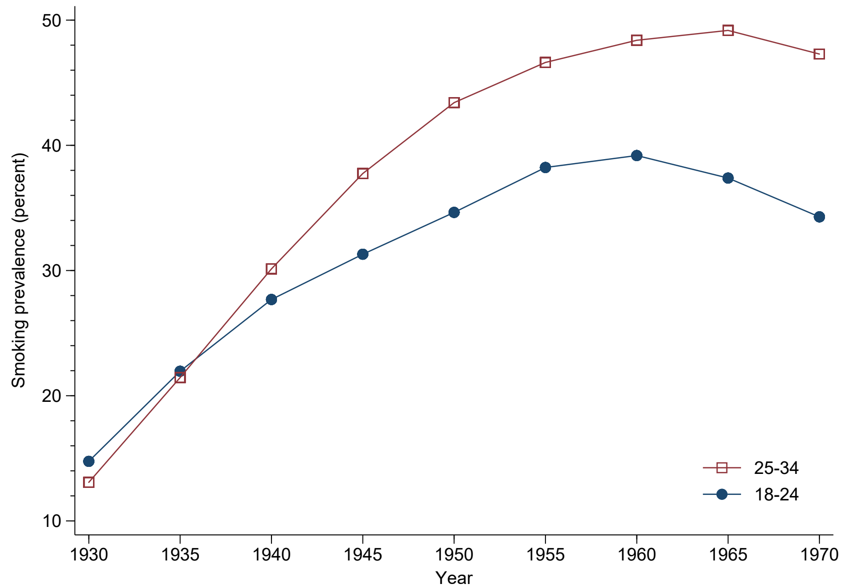


B: Population of each cohort at age 18



Panel A shows the number of births in the United States by year. Panel B shows the population of each cohort when they were age 18. Data comes from the Human Mortality Database, derived from vital statistics.

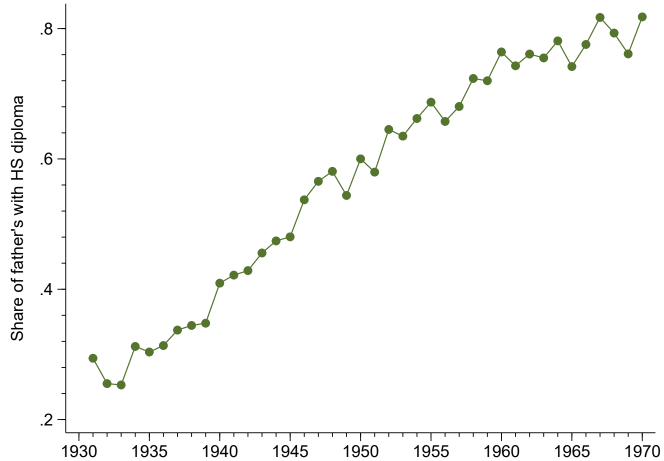
Appendix Figure 10: Estimates of smoking prevalence of American women of childbearing age



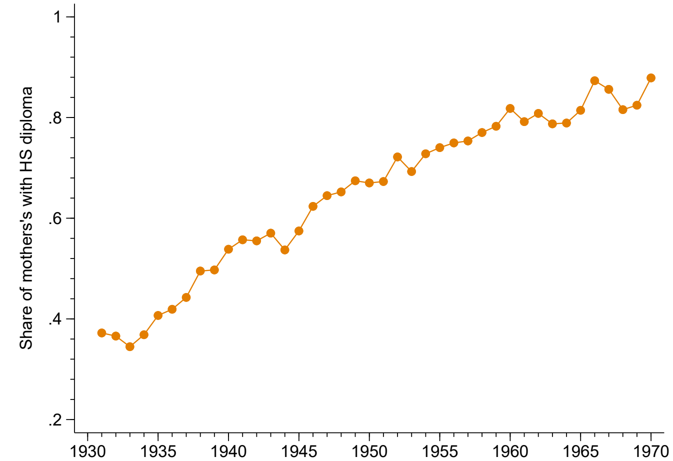
Based on estimates of smoking prevalence by age and year (every 5 years) from Holford et al. (2014), derived from survey data on retrospective smoking history. The figure plots age-adjusted smoking prevalence for women age 18-24 and 25-34 separately, assuming a uniform distribution of ages within age bins (ie. the unweighted average in each age bin across smoking rates by single year of age).

Appendix Figure 11: Parental education of White Americans by birth cohort

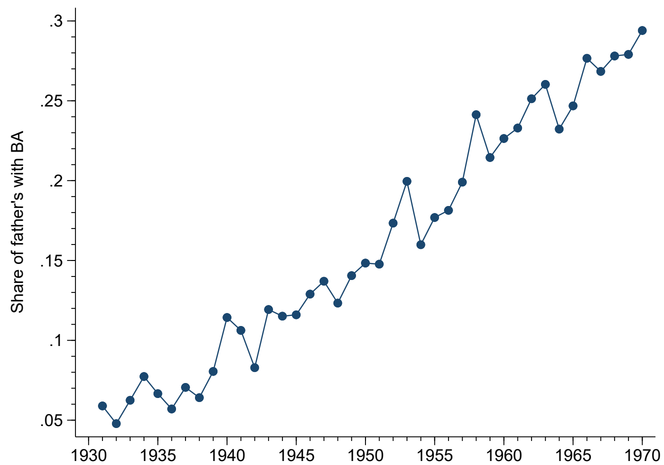
A: Share of fathers with high school diploma



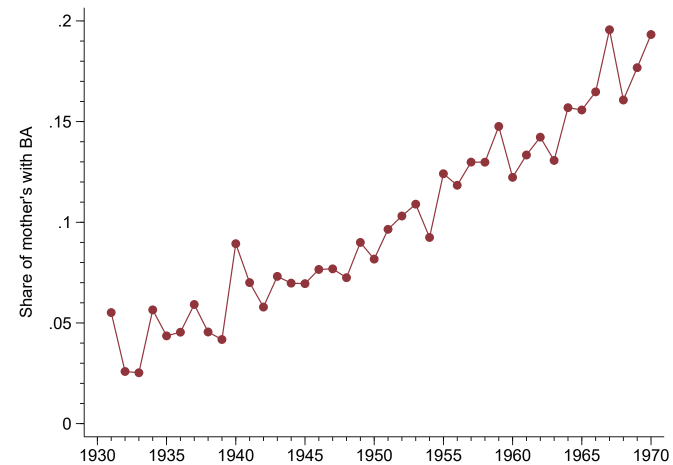
B: Share of mothers with high school diploma



A: Share of fathers with Bachelor's degree

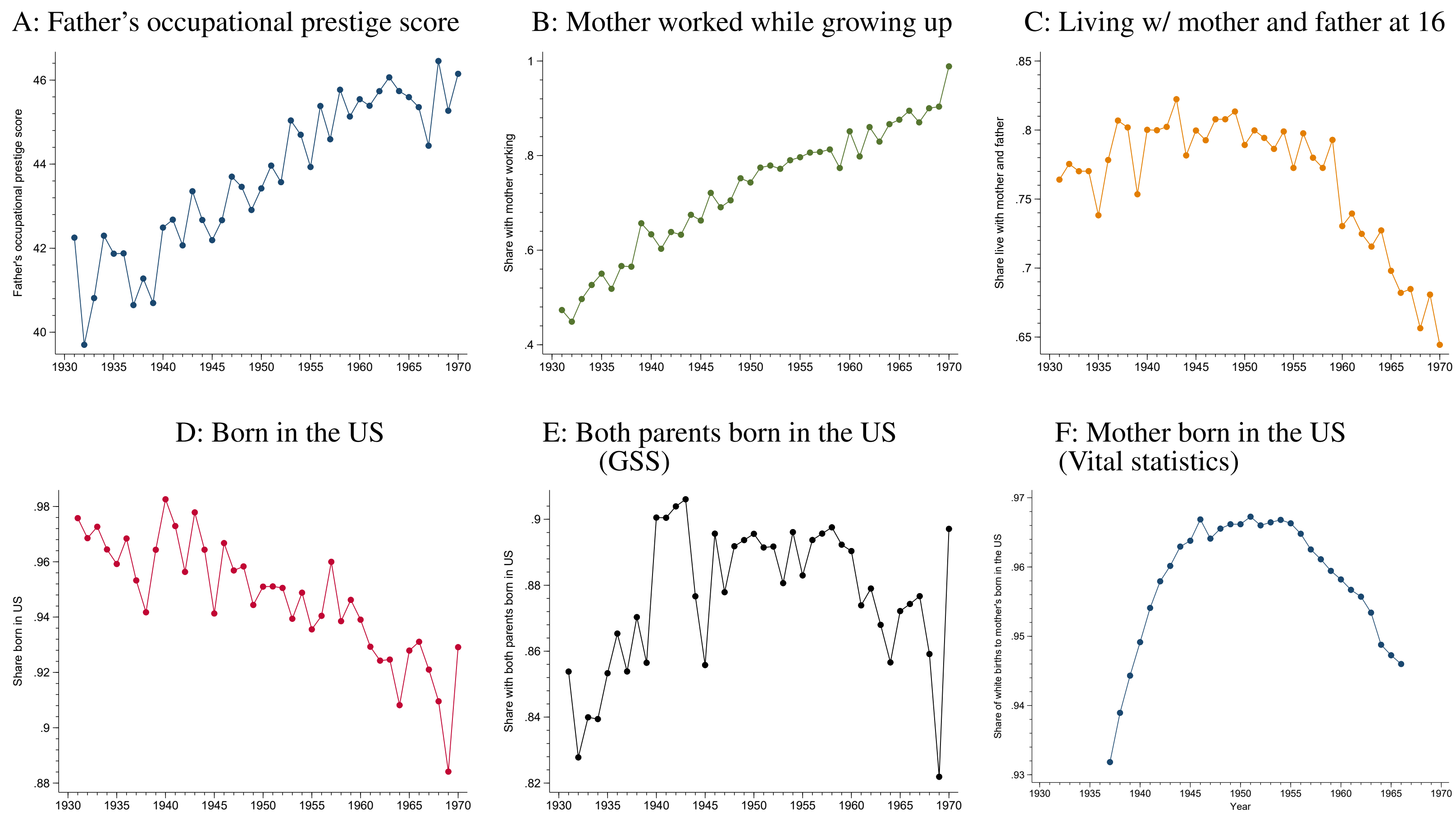


B: Share of mothers with Bachelor's degree



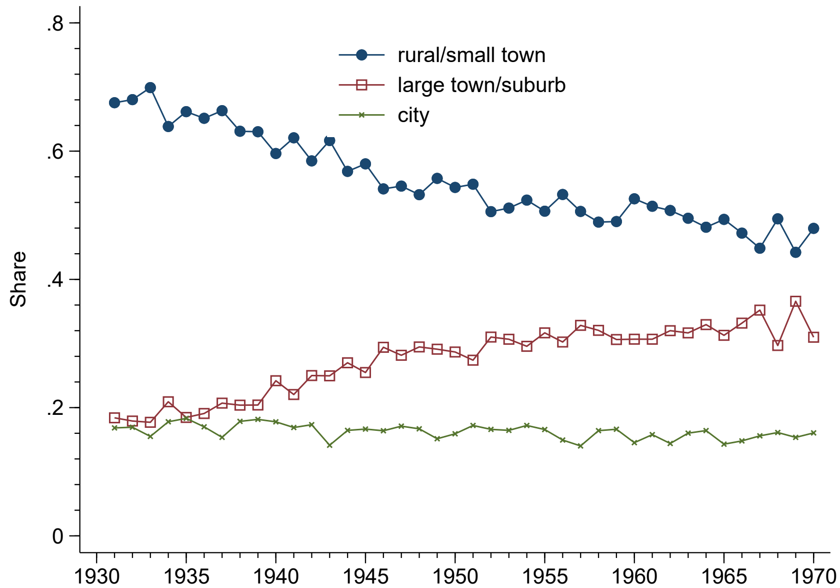
Each graph shows an estimate of parental educational attainment by individuals' birth cohort, estimated from the 1972-2016 waves of the General Social Survey. Each outcome is age-adjusted, by running a regression with cohort fixed effects and a quartic-in-age. The plots then show the estimated cohort effects, plus the estimated age effect for age 35. All regressions use sampling weights.

Appendix Figure 12: Family background and childhood circumstances of White Americans by birth cohort



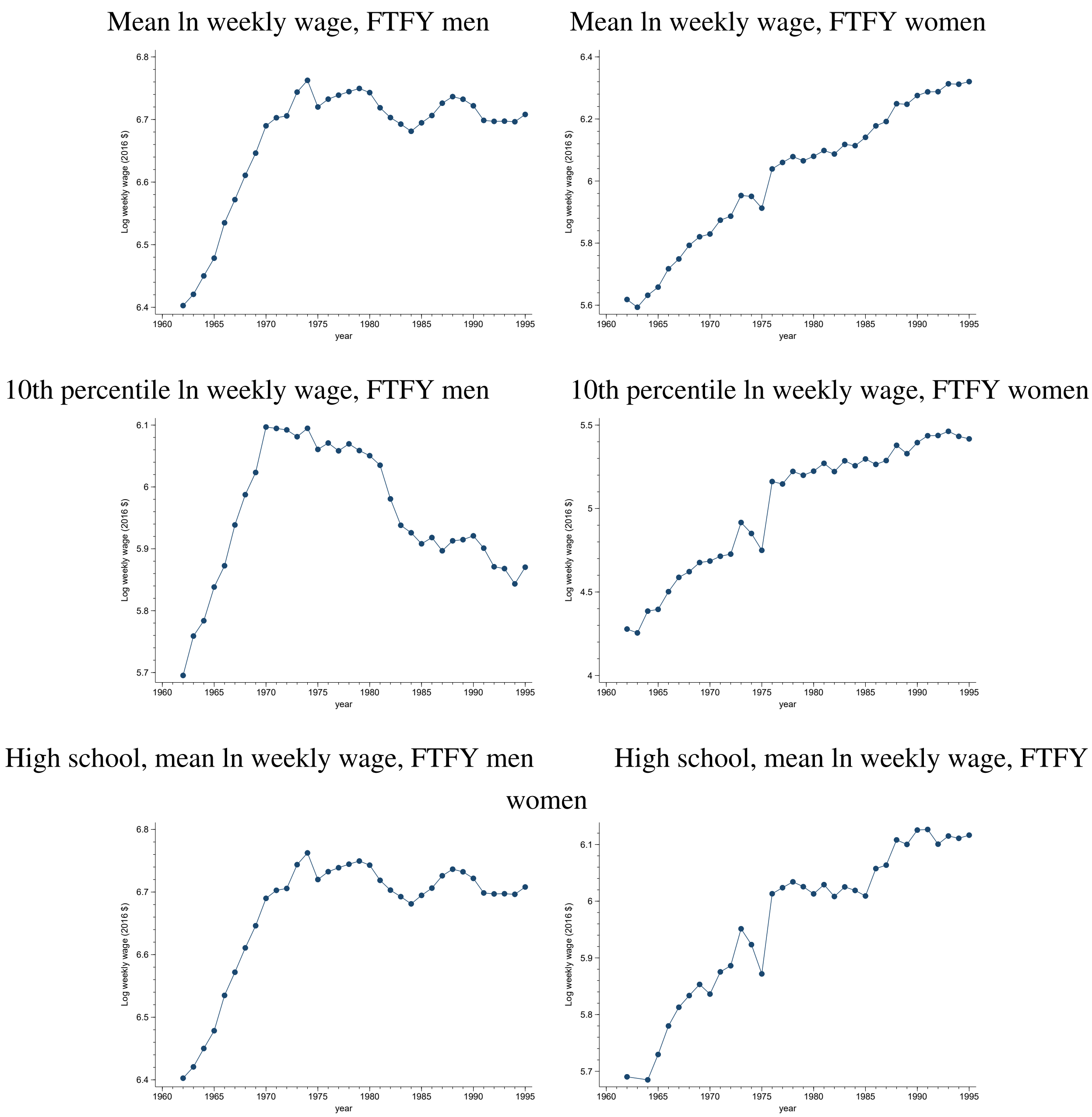
Panels A-E shows an estimate of the average value of the listed variable for white Americans by year of birth, estimated from the General Social Survey. Father's occupational prestige score is based on 1980 occupational classifications and is only available in 1988-2010, so Panel A is based on those years. The question on whether an individual's mother was working is only available 1994-2016, so panel B is based on only those years. Panel C is based on 1972-2016. Panel D and Panel E are based on 1977-2016. Each outcome is age-adjusted, by running a regression with cohort fixed effects and a quartic-in-age. The plots then show the estimated cohort effects, plus the estimated age effect for age 35. All regressions use sampling weights. Panel F reports estimates directly from vital statistics volumes which report parent's nativity.

Appendix Figure 13: Where white Americans lived at age 16, by birth cohort



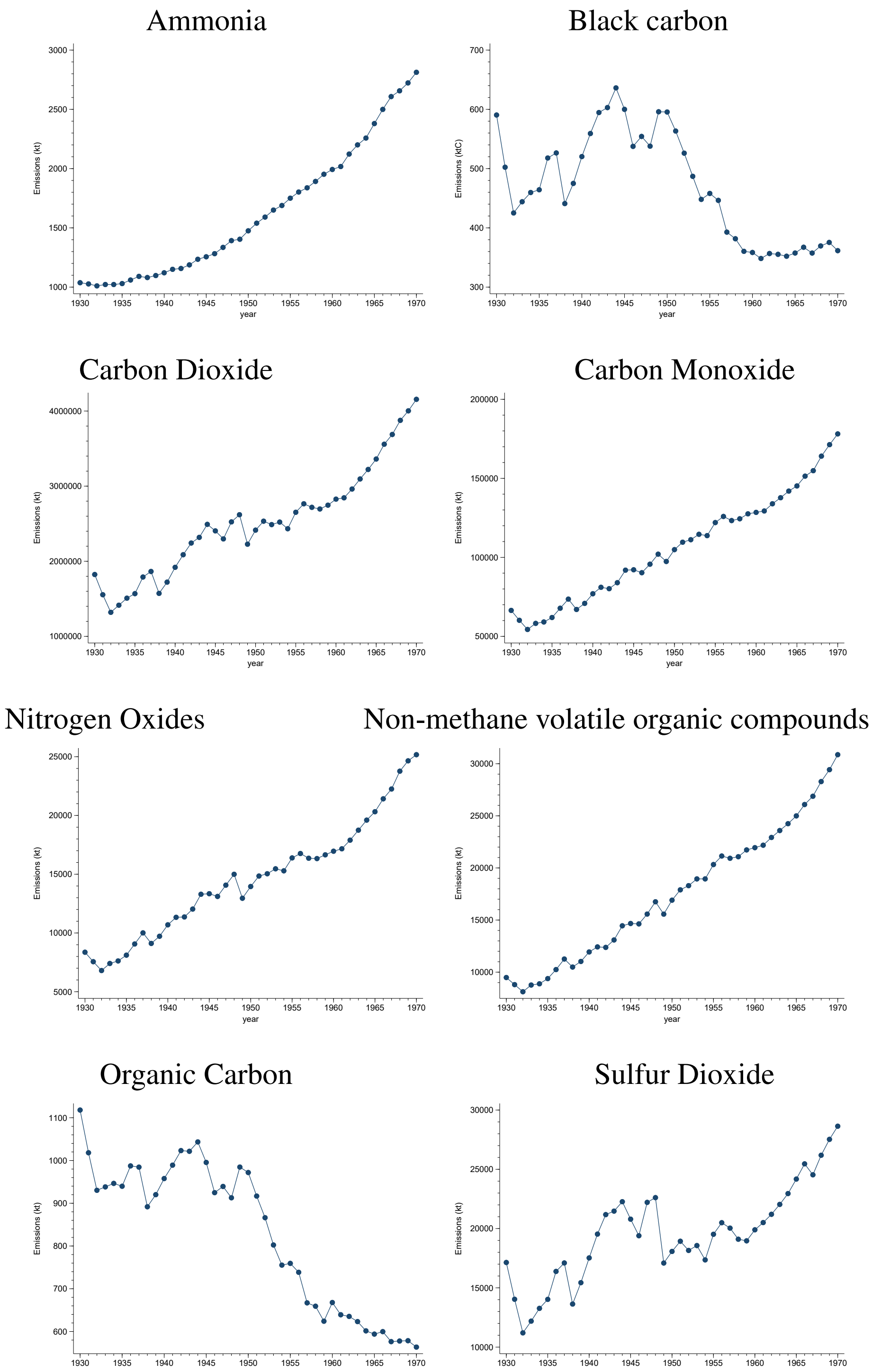
The figure shows estimates from the General Social Survey of the share of white Americans who lived in the listed type of place at age 16 by year of birth. "City" refers to large cities over 250,000 people. "Suburb" refers to a suburb near a large city. "Large town" refers to a city/town of 50,000 to 250,000. "rural/small town" includes smaller towns and rural areas. Each outcome is age-adjusted, by running a regression with cohort fixed effects and a quartic-in-age. The plots then show the estimated cohort effects, plus the estimated age effect for age 35. All regressions use sampling weights.

Appendix Figure 14: Wage trends, United States



Each panel shows the listed wage series estimated from the March Annual Social and Economic Supplement of the Current Population Survey (CPS).

Appendix Figure 15: Air pollution trends, United States



Each panel shows estimates from the Community Emissions Data System (O'Rourke et al.) of the trend in emissions of the listed air pollutant in the United States.

Appendix Table 1: Root mean squared error of year-specific Gompertz models of log mortality,
white men and women 1959-2014

	Within sample	Out of sample	
	ages 30-75	ages 20-29	ages 1-19
White men	.092	.703	.914
White women	.057	0.432	1.164

This table reports the root mean squared error of simple Gompertz regression models of log mortality for white men and women betweeen 1959 and 2019. Weighted least squares regressions of log mortality with a linear age term and a constant are estimated separately for each year, on observed log moratlity rates ages 30 to 75. The within sample column reports the within sample root mean squared residual of these models pooled across all years (separately for white men and white women). The out of sample columns report the root mean squared error when these models are extrapolated out of sample to younger ages, ages 20-29 and ages 1-19 separately.

Appendix Table 2: Shared cohort-specific trend break, log mortality of white Americans By Hispanic origin, 1997-2019

	(1) All whites	(2) Non-Hispanic whites	(3) Hispanic whites
Panel A: White women			
Average size of break	0.0165	0.0185	0.0149
	(0.0004)	(0.0004)	(0.0021)
Location of break	1948	1949	1939
	$[1948, 1949]$	$[1948, 1949]$	$[1938, 1945]$
P-value for existence of break	$< .001$	$< .001$	$< .001$
Panel B: White men			
Average size of break	0.0112	0.0131	-0.0043
	(0.0006)	(0.0006)	(0.0008)
Location of break	1946	1946	1958
	$[1944, 1946]$	$[1946, 1946]$	$[1957, 1960]$
P-value for existence of break	$< .001$	$< .001$	$< .001$
Linear age	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
Linear-age-by-year	Yes	Yes	Yes
Quadratic-age-by-year	No	No	No
Cubic-age-by-year	No	No	No

Each column shows the results of estimation of a model based on equation 4, with the log mortality rate for single age-by-year bins as the dependent variable. The columns respectively show results for the mortality rate of i) all whites, regardless of Hispanic origin; ii) non-Hispanic whites; and iii) Hispanic whites. All models are estimated by weighted least squares, following the approach outlined in Hansen (2000). The sample includes the years 1997-2019, ages 30-75, and cohorts born from 1930-1970. The row titled "Average size of break" reports the average value of $\delta_{2,c}$ across all years, with the standard error in parentheses calculated by the delta method. The row titled "Location of break" reports the estimated cohort at which a trend break occurs, with a 99 $\%$ confidence interval in brackets calculated by inverting the likelihood ratio statistic. The row titled "P-value for existence of break" reports p-value from an F-type test for the null hypothesis that no trend break occurs, based on 1000 bootstrap samples.

Appendix Table 3: Shared cohort-specific trend break, log mortality of white Americans By Census Region, 1980-2014

	(1) Northeast	(2) Midwest	(3) South	(4) West
Panel A: White women				
Average size of break	0.024	0.021	0.020	0.018
	(0.001)	(0.001)	(0.001)	(0.001)
Location of break	1949	1950	1948	1946
	[1949, 1949]	[1950, 1950]	[1947, 1948]	[1946, 1946]
P-value for existence of break	$< .001$	$< .001$	$< .001$	$< .001$
Panel B: White men				
Average size of break	0.026	0.029	0.026	0.026
	(0.001)	(0.001)	(0.001)	(0.001)
Location of break	1944	1946	1946	1942
	[1944, 1944]	[1946, 1946]	[1946, 1947]	[1942, 1943]
P-value for existence of break	$< .001$	$< .001$	$< .001$	$< .001$
Linear age	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes
Linear-age-by-year	Yes	Yes	Yes	Yes
Quadratic-age-by-year	No	No	No	No
Cubic-age-by-year	No	No	No	No

Each column shows the results of estimation of a model based on equation 4, with the log mortality rate of white men or women — in the listed Census Region — for single age-by-year bins as the dependent variable. All models are estimated by weighted least squares, following the approach outlined in Hansen (2000). The sample includes the years 1980-2014, ages 30-75, and cohorts born from 1930-1970. The row titled "Average size of break" reports the average value of $\delta_{2,c}$ across all years, with the standard error in parentheses calculated by the delta method. The row titled "Location of break" reports the estimated cohort at which a trend break occurs, with a 99 % confidence interval in brackets calculated by inverting the likelihood ratio statistic. The row titled "P-value for existence of break" reports p-value from an F-type test for the null hypothesis that no trend break occurs, based on 1000 bootstrap samples.

Appendix Table 4: Trend break estimates, Educational attainment of white Americans by birth cohort

		First break		Second break		
	Pre-trend	Location	Size	Location	Size	P-value
Years of schooling, men	0.0797	1947	-0.1110	1961	0.0743	$< .001$
	(.0019)	[1947, 1948]	(.0036)	[1958, 1963]	(.0059)	
Years of schooling, women	0.0816	1949	-0.0860	1963	0.0703	$< .001$
	(.0011)	[1948, 1949]	(.0023)	[1961, 1965]	(.0049)	
High school degree (or GED), men	0.0080	1948	-0.0090	1962	0.0031	$< .001$
	(.0001)	[1948, 1948]	(.0002)	[1959, 1965]	(.0004)	
High school degree (or GED), women	0.0080	1948	-0.0082	1964	0.0023	$< .001$
	(.0001)	[1947, 1948]	(.0002)	[1935, 1965]	(.0004)	
Some college, men	0.0122	1949	-0.0224	1960	0.0199	$< .001$
	(.0004)	[1948, 1949]	(.001)	[1958, 1963]	(.0015)	
Some college, women	0.0152	1949	-0.0147	1963	0.0121	$< .001$
	(.0003)	[1948, 1950]	(.0006)	[1959, 1965]	(.0013)	
Bachelor's degree, men	0.0079	1947	-0.0141	1960	0.0133	$< .001$
	(.0004)	[1946, 1948]	(.0008)	[1957, 1964]	(.0012)	
Bachelor's degree, women	0.0095	1950	-0.0108	1962	0.0137	$< .001$
	(.0002)	[1949, 1951]	(.0006)	[1959, 1965]	(.0011)	
Advanced degree, men	0.0041	1945	-0.0088	1960	0.0056	$< .001$
	(.0002)	[1945, 1946]	(.0003)	[1957, 1963]	(.0004)	
Advanced degree, women	0.0048	1948	-0.0076	1964	0.0101	$< .001$
	(.0002)	[1947, 1949]	(.0003)	[1961, 1965]	(.0008)	

Each row shows the estimation results of a separate trend break model which allow for two possible trend breaks of unknown location, with the listed dependent variable. All models are estimated using the sequential estimation approach suggested in Hansen (2000) for such models. The two columns titled "Location" reported the estimated location of the first and second trend breaks, respectively, with 99 % confidence intervals in brackets calculated by inverting the likelihood ratio statistic. The two columns titled "Size" report the magnitude of first and second trend breaks respectively, with standard errors in parentheses. The column titled "Pre-trend" reports the estimated trend prior to the first break. The column titled "P-value" reports the value of a bootstrap-based F-test suggested in Hansen (2000), for the null of a model with one break versus the alternative of a model with two breaks. I also conduct a similar test for the null of no break vs. the null of one break, which yields P-values $< .05$ for all variables.

I pool data from the CPS MORG data 1990 to 2018, white individuals age 25 to 75, cohort is defined as age - year -1. I then calculate approximate average years of schooling for each cohort based on the 16 schooling categories, and estimate the trend break models for average years of schooling by birth cohort.

Appendix Table 5: Trend break estimates, Birth order of white Americans by birth cohort

		First break		Second break		
	Pre-trend	Location	Size	Location	Size	P-value
Share 2nd born or later	-0.0095	1941	0.0176	1960	-0.0219	$<$ .001
	(.0013)	[1937, 1943]	(.0017)	[1958, 1962]	(.0017)	
Share 3rd born or later	-0.0083	1946	0.0185	1963	-0.0368	$<$ .001
	(.0008)	[1942, 1947]	(.0013)	[1961, 1964]	(.0025)	
Share 4th born or later	-0.0086	1947	0.0156	1964	-0.0288	$<$ .001
	(.0004)	[1946, 1948]	(.0007)	[1962, 1965]	(.0017)	
Share 5th born or later	-0.0068	1947	0.0118	1964	-0.0210	$<$ .001
	(.0003)	[1946, 1948]	(.0006)	[1962, 1965]	(.0013)	
Simulated effect on ln(mort), men	-0.0044	1947	0.0073	1964	-0.0129	$<$ .001
	(.0002)	[1946, 1948]	(.0004)	[1962, 1965]	(.0009)	
Simulated effect on ln(mort), men	-0.0061	1947	0.0107	1964	-0.0201	$<$ .001
	(.0003)	[1946, 1948]	(.0006)	[1962, 1965]	(.0013)	

Each row shows the estimation results of a separate trend break model which allow for two possible trend breaks of unknown location, with the listed dependent variable. All models are estimated using the sequential estimation approach suggested in Hansen (2000) for such models. The two columns titled "Location" reported the estimated location of the first and second trend breaks, respectively, with 99 % confidence intervals in brackets calculated by inverting the likelihood ratio statistic. The two columns titled "Size" report the magnitude of first and second trend breaks respectively, with standard errors in parentheses. The column titled "Pre-trend" reports the estimated trend prior to the first break. The column titled "P-value" reports the value of a bootstrap-based F-test suggested in Hansen (2000), for the null of a model with one break versus the alternative of a model with two breaks. I also conduct a similar test for the null of no break vs. the null of one break, which yields P-values $< .05$ for all variables.

Observed birth order shares for white Americans comes from Vital Statistics volumes and Heuser (1976). The simulated effect of birth order on log mortality is derived from within-family estimates of the effect of birth order on mortality from Barclay and Kolk (2015) and observed birth order shares for white Americans. I then estimate the described trend break models by birth cohort on the simulated data.

Appendix Table 6: Trend break estimates, Cohort size

		First break		Second break		
	Pre-trend	Location	Size	Location	Size	P-value
Cohort size at age 18 (mil.)	-0.026	1936	0.139	1957	-0.167	$< .001$
	(.016)	[1935, 1939]	(.018)	[1956, 1959]	(.009)	
Cohort size at birth (mil.)	0.091	1958	-0.169			$< .001$
	(.004)	[1956, 1960]	(.011)			

Each row shows the estimation results of a separate trend break model which allow for two possible trend breaks of unknown location, with the listed dependent variable. All models are estimated using the sequential estimation approach suggested in Hansen (2000) for such models. The two columns titled "Location" reported the estimated location of the first and second trend breaks, respectively, with 99 $\%$ confidence intervals in brackets calculated by inverting the likelihood ratio statistic. The two columns titled "Size" report the magnitude of first and second trend breaks respectively, with standard errors in parentheses. The column titled "Pre-trend" reports the estimated trend prior to the first break. I conduct bootstrap-based F-tests suggested in Hansen (2000), for i) the null of a model with one break versus the alternative of a model with two breaks, and ii) for the null of no break vs. the null of one break. When the p-value for i) is $< .05$ then I report results from the model with two breaks, and the column titled "P-value" reports the p-value from i). When the p-value for i) is $\geq .05$, I report results from the model with one break, and the p-value is that from test ii).

All data is from the Human Mortality Database.

Appendix Table 7: Trend break estimates, Smoking of American women of childbearing age, by year

	Pre-trend	Location	Size	P-value
Smoking prev., women 18-35	1.216	1955	-1.401	$< .001$
(percent)	(.059)	[1945, 1955]	(.145)	

This table shows the estimation results of a trend break model which allow for two possible trend breaks of unknown location, with the listed dependent variable. It is estimated using the sequential estimation approach suggested in Hansen (2000) for such models. I conduct bootstrap-based F-tests suggested in Hansen (2000), for i) the null of a model with one break versus the alternative of a model with two breaks, and ii) for the null of no break vs. the null of one break. The p-value for i) is $\geq .05$, so I report results from the model with one break, and the value in the column labelled "P-value" is the p-value from test ii). The column titled "Location" reported the estimated location of the trend break, with 99 $\%$ confidence intervals in brackets calculated by inverting the likelihood ratio statistic. The column titled "Size" report the magnitude of second trend breaks respectively, with standard errors in parentheses. The column titled "Pre-trend" reports the estimated trend prior to the first break.

Based on estimates of smoking prevalence by age and year (every 5 years) from Holford et al. (2014), derived from survey data on retrospective smoking history. I fit the trend break models to prevalence estimates for women ages 18-35.

Appendix Table 8: Trend break estimates, Family background and childhood circumstances of white Americans, by cohort

		First break		Second break		
	Pre-trend	Location	Size	Location	Size	P-value
Share whose father has BA	-0.002	1936	0.008			$<$ .001
	(.002)	[1935, 1951]	(.002)			
Share whose mother has BA	0.003	1948	0.002			0.001
	(.0003)	[1935, 1965]	(.0006)			
Share whose father has HS diploma	0.005	1935	0.014	1955	-0.011	0.029
	(.004)	[1935, 1965]	(.004)	[1948, 1961]	(.001)	
Share whose mother has HS diploma	0.017	1952	-0.009			$<$ .001
	(.0006)	[1945, 1959]	(.001)			
Father's occupational prestige	0.154	1962	-0.151			0.103
	(.012)	[1935, 1965]	(.068)			
Mother worked while child growing up	0.010	1965	-0.011			0.073
	(.0004)	[1935, 1965]	(.004)			
Living w/ mother and father at 16	0.001	1954	-0.011			$<$ .001
	(.0004)	[1949, 1958]	(.0009)			
Born in the US	-0.001	1957	-0.002			0.022
	(.0002)	[1935, 1965]	(.0007)			
Both parents born in US	0.007	1942	-0.008			$<$ .001
	(.0008)	[1938, 1951]	(.001)			
Mother born in the US	0.004	1944	-0.004	1956	-0.002	$<$ .001
(Vital statistics)	(0.0002)	[1943, 1944]	(0.0002)	[1953, 1961]	(0.0002)	
Lived in rural/small town when 16	-0.008	1948	0.005			$<$ .001
	(.0007)	[1942, 1959]	(.001)			
Lived in big town/suburb when 16	0.007	1949	-0.005			$<$ .001
	(.0005)	[1944, 1958]	(.0009)			
Lived in city when 16	-0.002	1965	0.002			0.681
	(.0002)	[1935, 1965]	(.002)			

Each row shows the estimation results of a separate trend break model which allow for two possible trend breaks of unknown location, with the listed dependent variable. All models are estimated using the sequential estimation approach suggested in Hansen (2000) for such models. The two columns titled "Location" reported the estimated location of the first and second trend breaks, respectively, with 99 % confidence intervals in brackets calculated by inverting the likelihood ratio statistic. The two columns titled "Size" report the magnitude of first and second trend breaks respectively, with standard errors in parentheses. The column titled "Pre-trend" reports the estimated trend prior to the first break. I conduct bootstrap-based F-tests suggested in Hansen (2000), for i) the null of a model with one break versus the alternative of a model with two breaks, and ii) for the null of no break vs. the null of one break. When the p-value for i) is $<$ .05 then I report results from the model with two breaks, and the column titled "P-value" reports the p-value from i). When the p-value for i) is $\geq$ .05, I report results from the model with one break, and the p-value is that from test ii).

With one exception, all data come from various waves of the General Social Survey. shows an estimate of the average value of the listed variable for white Americans by year of birth, estimated from the General Social Survey. I first age-adjusted each outcome, by running a regression with cohort fixed effects and a quartic-in-age. I then run the trend break models on these age-adjusted series by cohort (which are the estimated cohort effects, plus the estimated age effect for age 35). See the notes to the corresponding Appendix Figures for more detail on exact GSS waves for each variable. The data for "Mother born in the US" come from vital statistics volumes which report the number of births in a year by parent's nativity.

Appendix Table 9: Trend break estimates, Wages in the United States, by year

	Pre-trend	Location	Size	P-value
Mean ln wage, men	0.035	1972	-0.037	$< .001$
	(.001)	[1971, 1973]	(.002)	
10th pct ln wage, men	0.047	1971	-0.058	$< .001$
	(.002)	[1970, 1971]	(.0024)	
Mean ln wage, women	0.030	1976	-0.013	$< .001$
	(.001)	[1971, 1979]	(.002)	
10th pct ln wage, women	0.056	1979	-0.040	$< .001$
	(.002)	[1977, 1982]	(.004)	
Mean ln wage, HS-only men	0.035	1972	-0.037	$< .001$
	(.001)	[1971, 1973]	(.002)	
Mean ln wage, HS-only women	0.021	1977	-0.014	$< .001$
	(.001)	[1972, 1980]	(.002)	

Each row shows the estimation results of a separate trend break model which allow for two possible trend breaks of unknown location, with the listed dependent variable. All models are estimated using the sequential estimation approach suggested in Hansen (2000) for such models. I conduct bootstrap-based F-tests suggested in Hansen (2000), for i) the null of a model with one break versus the alternative of a model with two breaks, and ii) for the null of no break vs. the null of one break. For all dependent variables, the p-value for i) is $\geq .05$, so I report results from the model with one break, and the value in the column labelled "P-value" is the p-value from test ii). The column titled "Location" report the estimated location of the trend break, with 99 % confidence intervals in brackets calculated by inverting the likelihood ratio statistic. The column titled "Size" report the magnitude of second trend breaks respectively, with standard errors in parentheses. The column titled "Pre-trend" reports the estimated trend prior to the first break.

I calculate each wage series from the March Annual Social and Economic Supplement of the Current Population Survey (CPS). I then fit the trend break models to the annual time series.

Appendix Table 10: Trend break estimates, Emissions in the United States by year

		First break		Second break		
	Pre-trend	Location	Size	Location	Size	P-value
Black carbon (kt)	6.21	1949	-27.4	1961	26.018	0.001
	(1.28)	[1944, 1951]	(2.87)	[1935, 1965]	(4.76)	
Carbon monoxide (kt)	593	1935	2,130	1965	3,815	0.005
	(503)	[1935, 1940]	(532)	[1963, 1965]	(532)	
Carbon dioxide (kt)	64,555	1945	-34,990	1961	119,7063	0.030
	(6089)	[1936, 1965]	(10024)	[1957, 1964]	(15223)	
Ammonia (kt)	14.0	1944	36.8	1962	41.4	$< .001$
	(1.01)	[1942, 1945]	(1.53)	[1960, 1964]	(2.52)	
Non-methane volatile organic compounds (kt)	114	1935	403	1965	655	$< .001$
	(73.5)	[1935, 1940]	(77.8)	[1963, 1965]	(77.8)	
Nitrogen oxides (kt)	417	1962	702	1947	-154	0.005
	(21.3)	[1960, 1964]	(71.0)	[1938, 1958]	(39.1)	
Organic carbon (kt)	-28.3	1943	-31.6	1935	40.9	0.025
	(8.99)	[1941, 1947]	(4.48)	[1935, 1965]	(11.8)	
Sulfur dioxide (kt)	628	1959	949	1943	-720	$< .001$
	(87.1)	[1952, 1963]	(152)	[1939, 1948]	(133)	

Each row shows the estimation results of a separate trend break model which allow for two possible trend breaks of unknown location, with the listed dependent variable. All models are estimated using the sequential estimation approach suggested in Hansen (2000) for such models. The two columns titled “Location” reported the estimated location of the first and second trend breaks, respectively, with 99 % confidence intervals in brackets calculated by inverting the likelihood ratio statistic. The two columns titled “Size” report the magnitude of first and second trend breaks respectively, with standard errors in parentheses. The column titled “P-value” reports the value of a bootstrap-based F-test suggested in Hansen (2000), for the null of a model with one break versus the alternative of a model with two breaks. I also conduct a similar test for the null of no break vs. the null of one break, which yields P-values $< .05$ for all variables.

All data is an estimate of the time series in emissions of the listed air pollutant in the United States, from the Community Emissions Data System (O’Rourke et al.).

References

E Arias and J Xu. United states life tables, 2019. *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 70(19):1–59, 2022.

Kieron Barclay and Martin Kolk. Birth order and mortality: a population-based cohort study. *Demography*, 52(2):613–639, 2015.

Bruce E Hansen. Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603, 2000.

Robert L Heuser. *Fertility tables for birth cohorts by color: United States, 1917-73*. US Department of Health, Education, and Welfare, Public Health Service , 1976.

P. R O'Rourke, S. J. Smith, A. Mott, H. Ahsan, E. E. McDuffie, M. Crippa, S. Klimont, B. McDonald, Z. Wang, M. B. Nicholson, L. Feng, and R. M. Hoesly. Ceds v-2021-02-05 emission data 1975-2019 (2021, february 05). *Zenodo, http://doi.org/10.5281/zenodo.4509372.*

Hannes Schwandt and Till M Von Wachter. Socioeconomic decline and death: Midlife impacts of graduating in a recession. Technical report, National Bureau of Economic Research, 2020.

Cyrus Shahpar and Guohua Li. Homicide mortality in the united states, 1935–1994: age, period, and cohort effects. *American journal of epidemiology*, 150(11):1213–1222, 1999.

US Census Bureau. *Historical statistics of the United States, colonial times to 1970*. Number 78. US Department of Commerce, Bureau of the Census, 1975.