Childhood Health, Sibling Outcomes and the 1918 Influenza Pandemic

John Parman
College of William & Mary and NBER
July 11, 2013

State Gym at Iowa State University; Image Special Collections Department, ISU Library

Motivation

- There is an increasing body of evidence that poor health as a child has lasting impacts on educational and labor market outcomes
- A portion of this relationship is driven by biology, with poor infant and childhood health being correlated with adult health complications
- However, there is also an economic channel: poor childhood health changes the relative prices faced by parents when investing in children
- Raises the possibility that even healthy children will have their outcomes affected by unhealthy siblings

Brief overview of paper

- I create a new dataset linking adult health and educational attainment to childhood household characteristics
- I estimate the effects of having a sibling in utero during the pandemic on educational and health outcomes
- Results suggest that parents shifted resources to older children when a child received a negative health shock
- The effects are large and suggest health shocks or interventions can dramatically alter the relative outcomes of siblings

Childhood Health and Adult Outcomes

- There is a growing body of empirical evidence linking childhood health to adult socioeconomic and health outcomes
- One strand of the literature looks at health differences between twins or siblings to control for household fixed effects
- Another strand focuses on shocks to childhood health that target specific birth cohorts
- The studies generally find that poor health in early childhood is associated with poor economic and health outcomes as an adult

Childhood Health and Adult Outcomes

Possible mechanisms for the link between childhood health and adult outcomes:

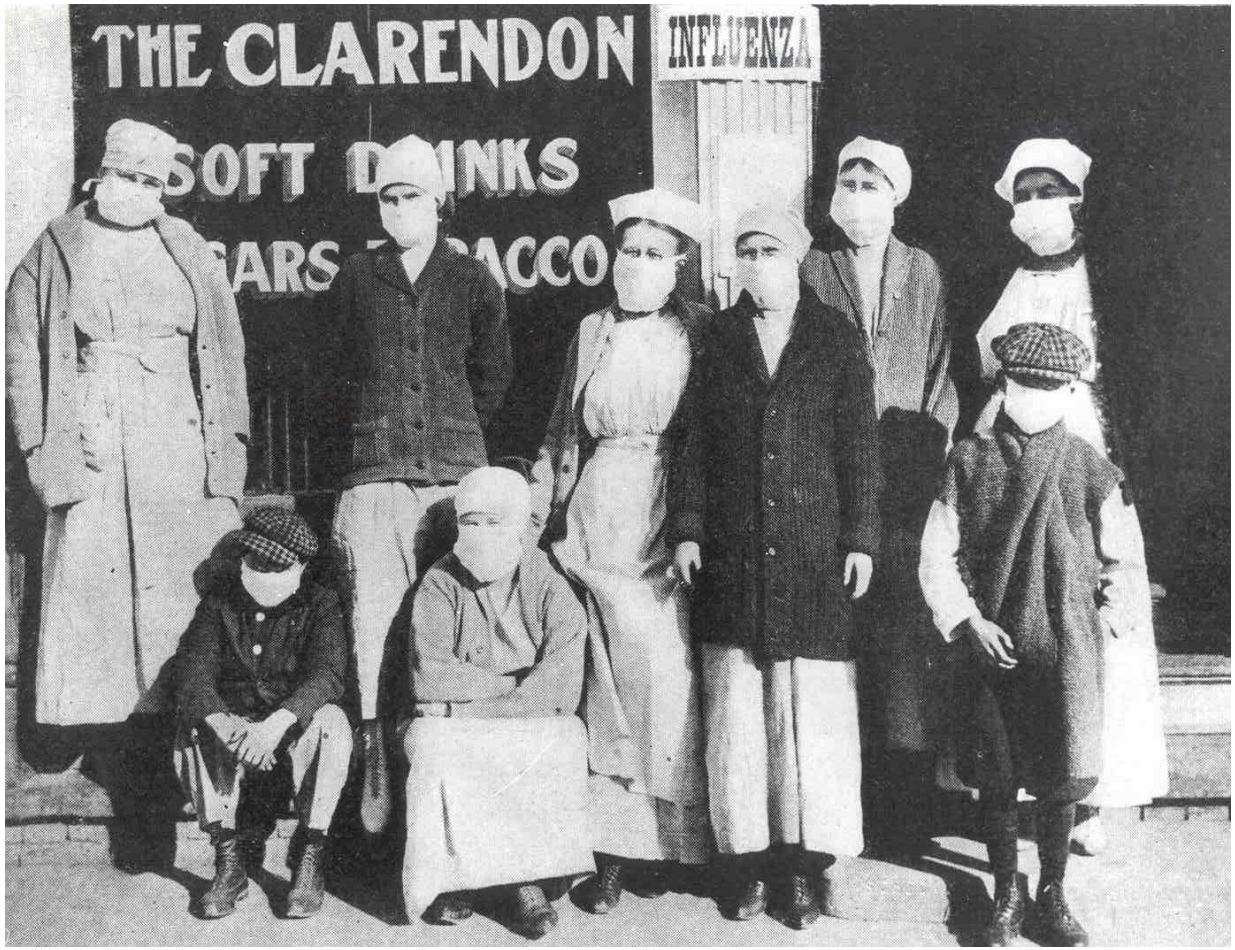
- Direct effect of poor childhood health on school attendance and productivity
- Poor health hindering cognitive development
- In utero health affecting metabolism leading to later chronic health issues
- Parents altering investment decisions based on children's health

Childhood Health and Adult Outcomes

Parents altering investment decisions implies sibling outcomes may be affected:

- Becker and Tomes framework predicts increased investment in less healthy child if health shock affects endowment but not price of increasing child quality
- Investment will reinforce sibling inequality if health shock leads to higher price of increasing child quality
- More recent models (Behrman et al., Ejrnæs & Pörtner, Adhvaryu & Nyshadham) incorporate aversion to inequality in outcomes across children

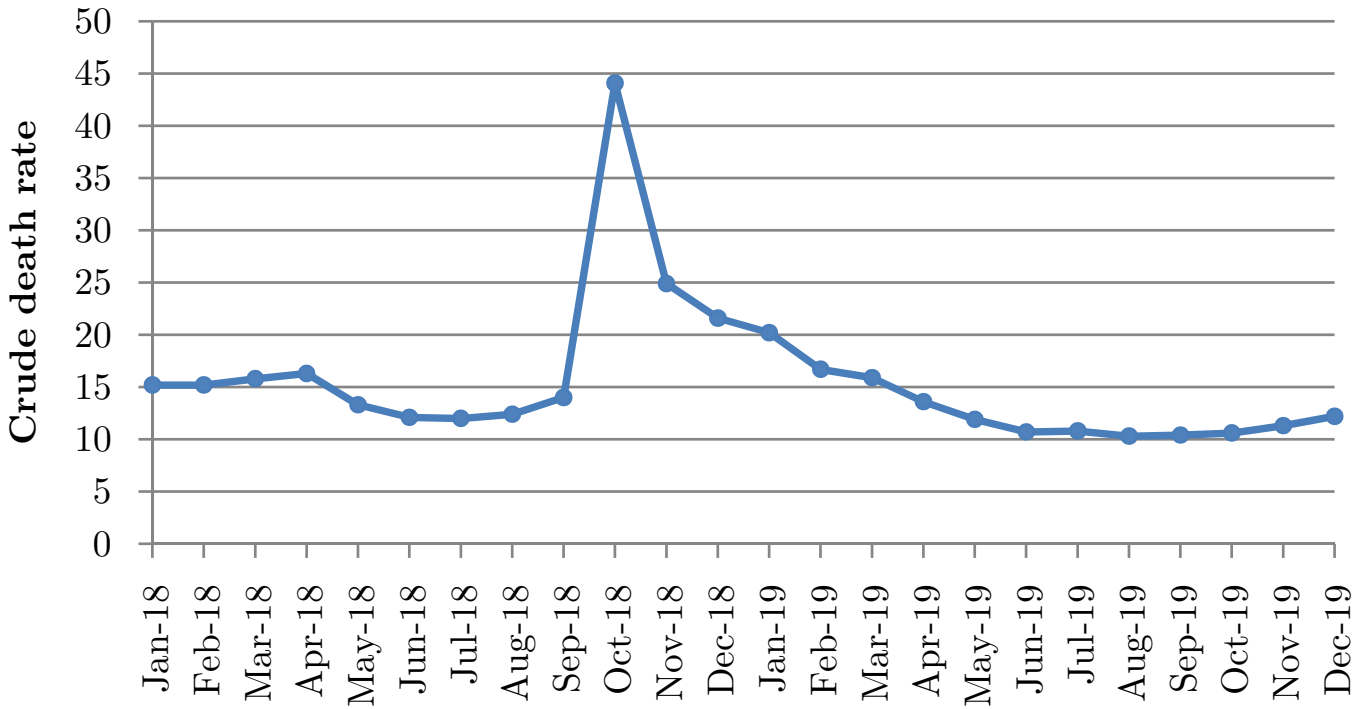
The Influenza Pandemic of 1918



Credit: Office of the Public Health Service Historian (image at www.flu.gov)

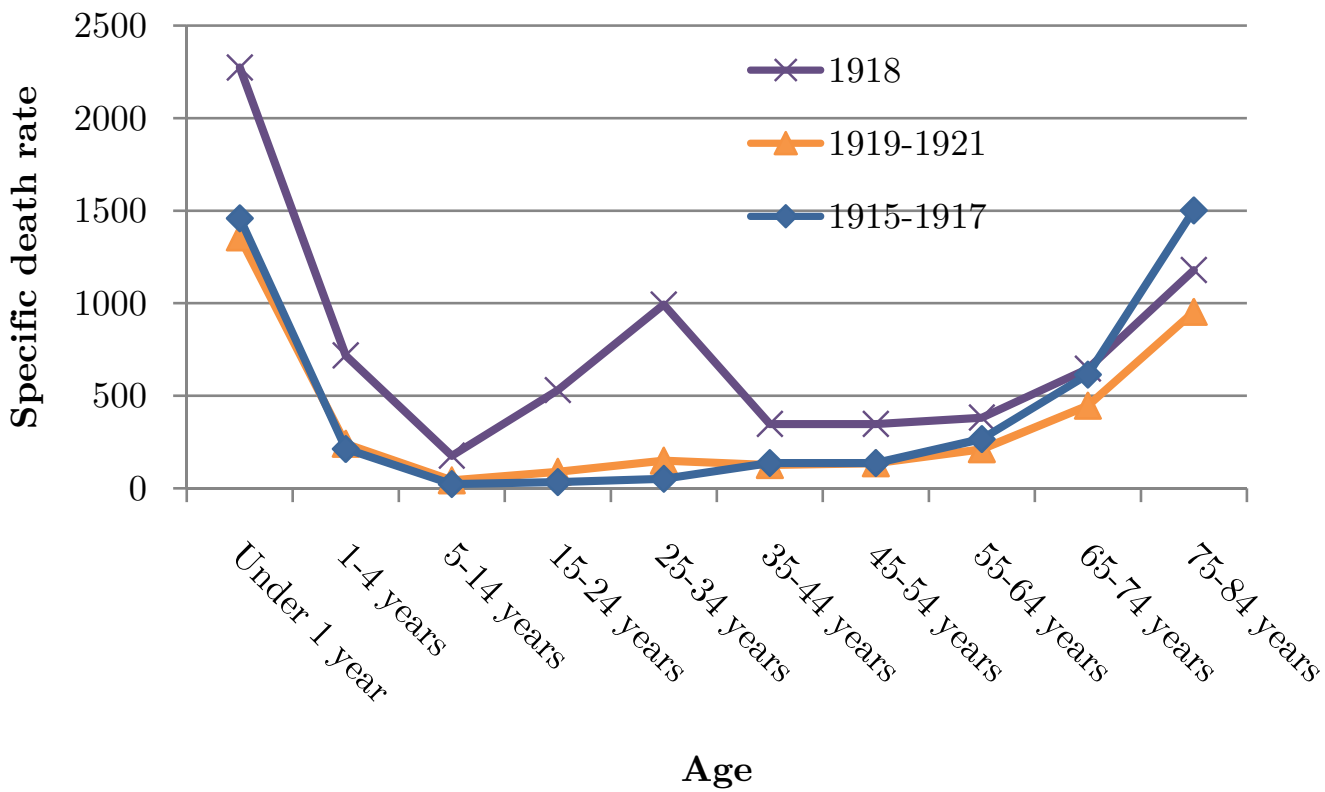
The Timing of the Pandemic

The Timing of the Pandemic



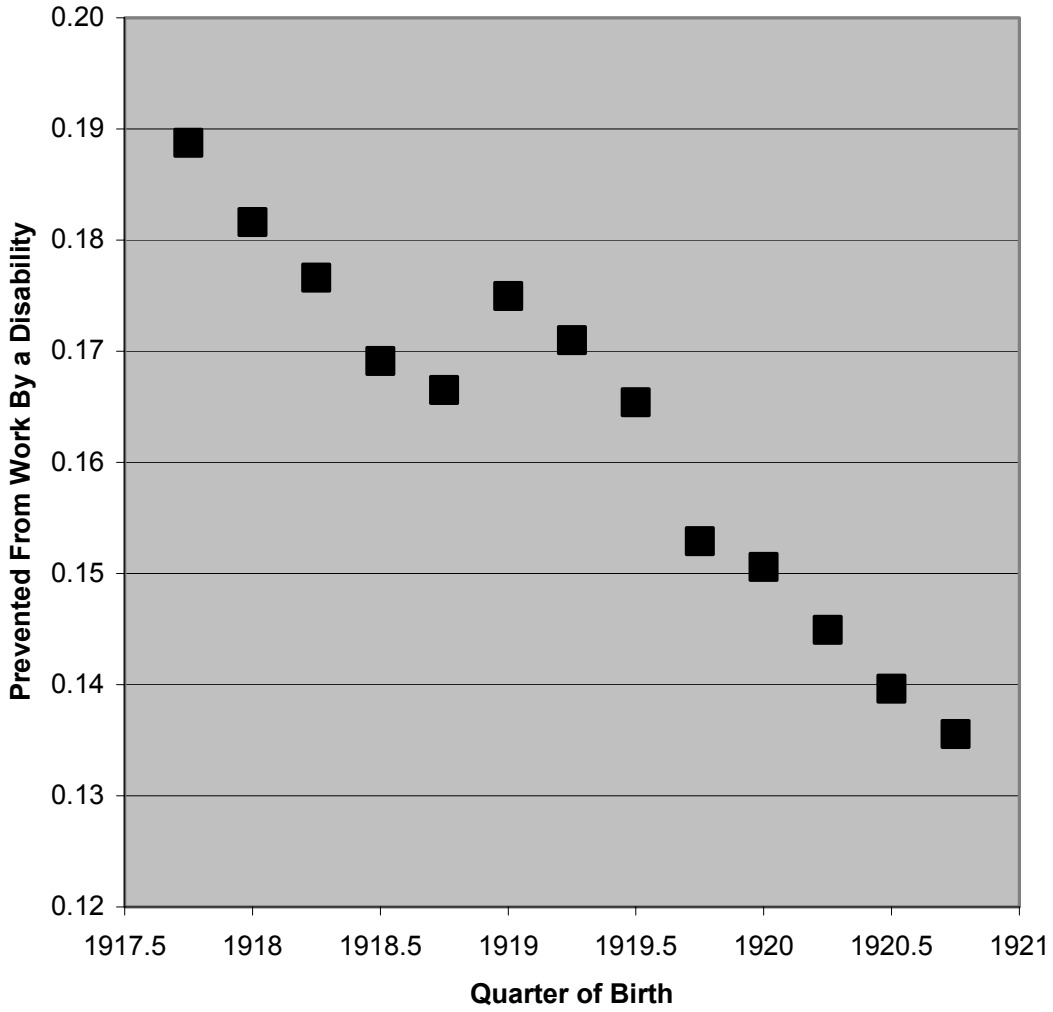
Deaths per 1,000 people by month, 1918-1919

The Distribution of Flu Deaths



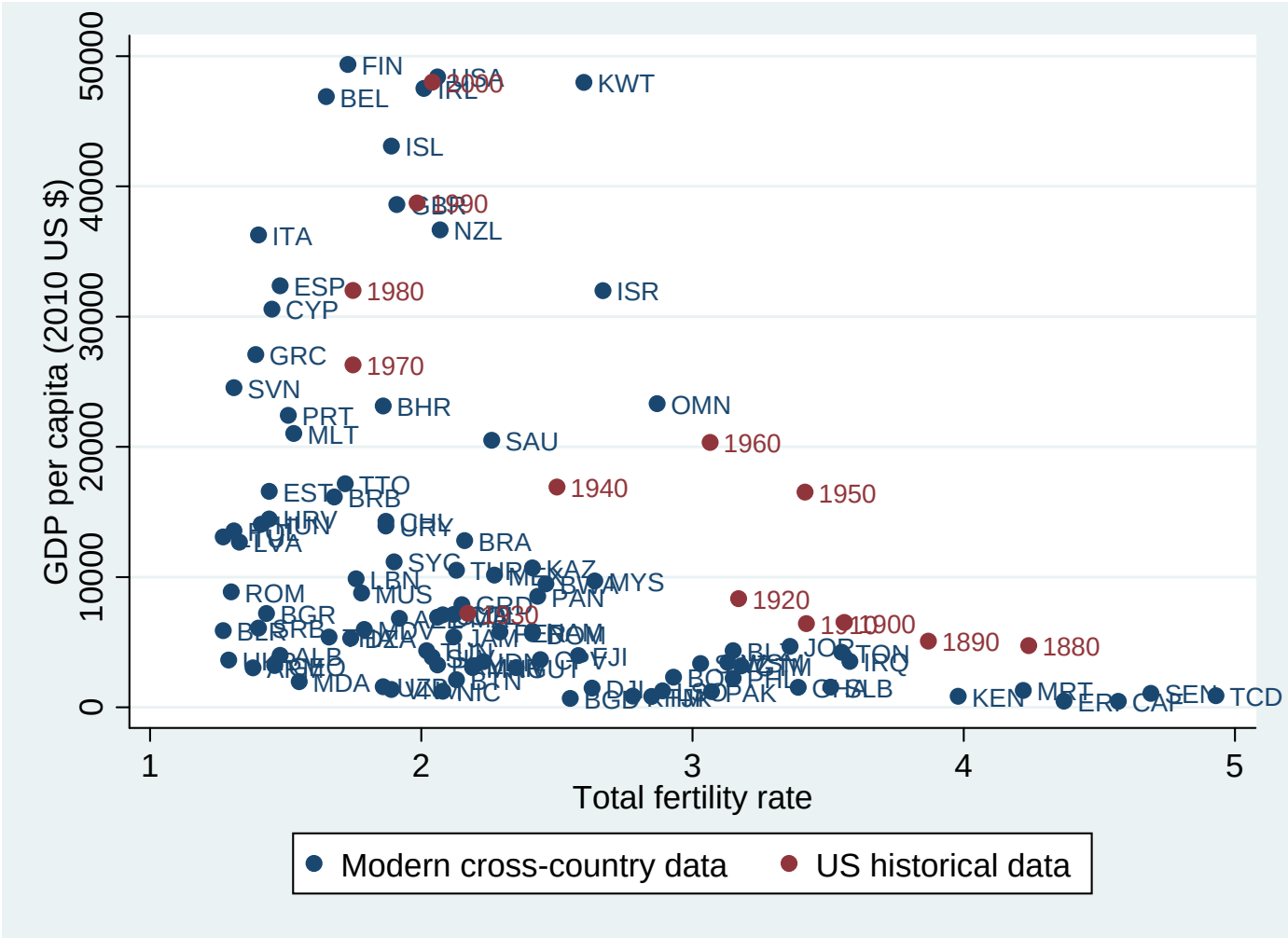
Deaths from influenza or pneumonia per 100,000 people by age group

The Distribution of Flu Deaths



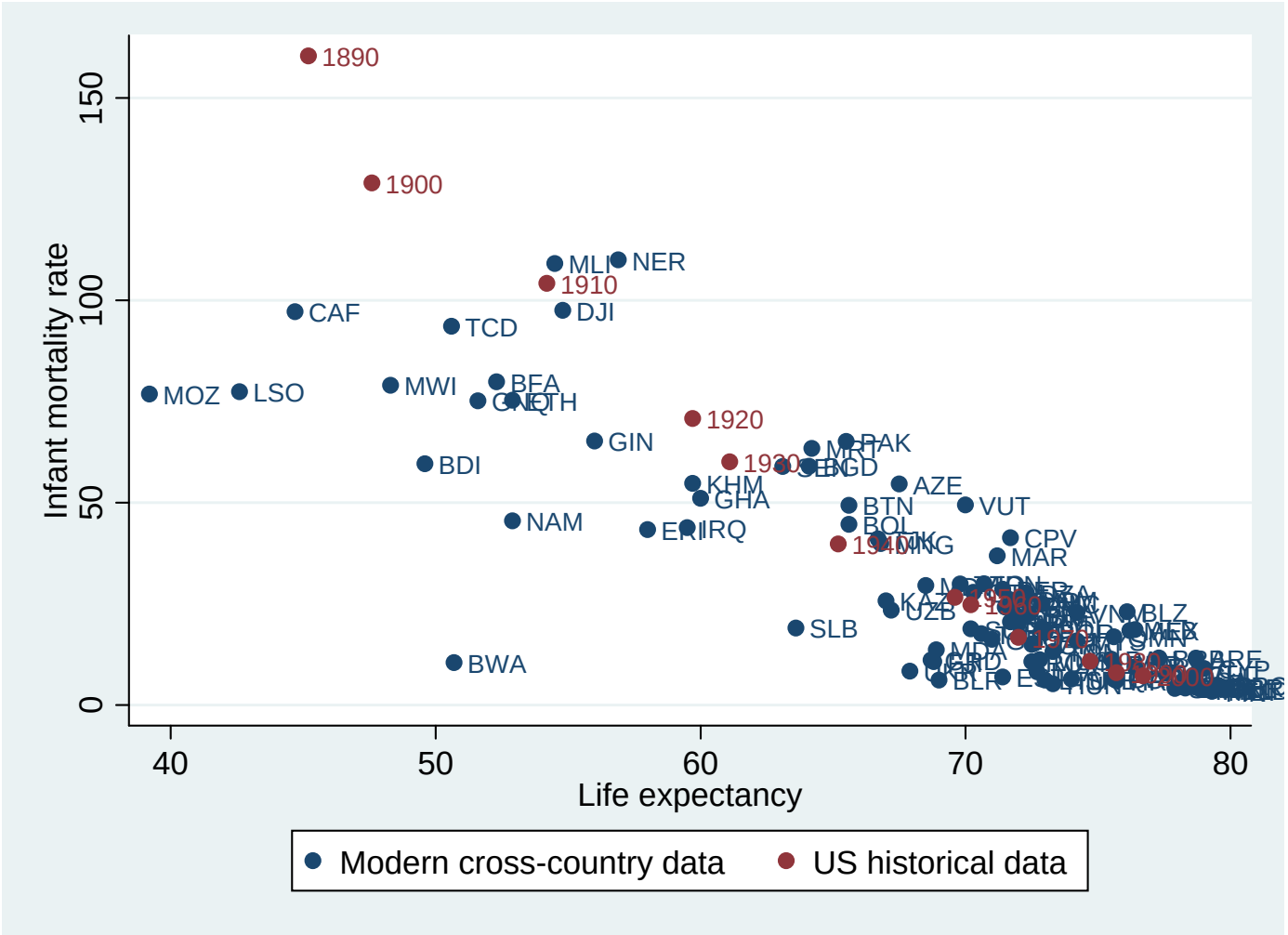
1980 male disability rate, Figure 7 from Almond (2006)

Placing the Pandemic in Context



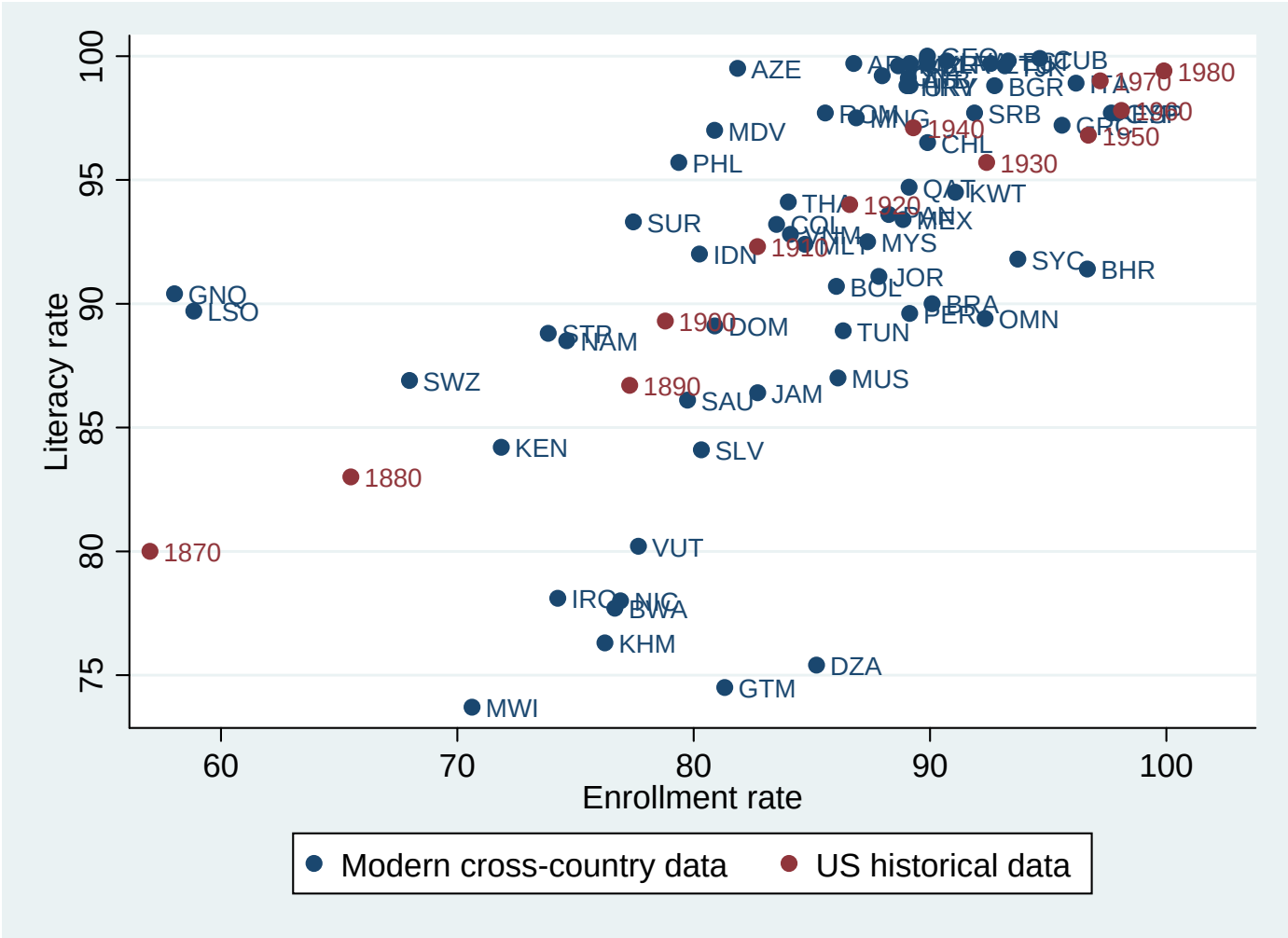
GDP per capita and total fertility rate by country for 2010 and by decade for the US

Placing the Pandemic in Context



Infant mortality rates and life expectancy by country for 2010 and by decade for the US

Placing the Pandemic in Context



Literacy rates and enrollment rates by country for 2010 and by decade for the US

The Influenza Pandemic of 1918 as a Natural Experiment

- It arrived rapidly and unexpectedly, families weren't making decisions in anticipation of the pandemic
- It was a massive health shock striking a quarter of the population and killing over 600,000 Americans but was short lived
- The flu had particularly severe effects for pregnant women that led to poor health outcomes for children in utero during the pandemic
- US fertility, health and education measures at the time of the pandemic are comparable to those of modern developing countries

Potential Data Sources

- The basic idea is to look at how adult outcomes depend on having a sibling affected by the influenza pandemic
- This requires person-level data with adult outcomes (ideally income, occupation, education, health) that also have info on siblings
- Almond uses modern 1960/1970/1980 census data, Almond & Mazumder use SIPP data but these lack sibling info
- Historical census data allows for observing siblings but not adult outcomes
- Solution: combine childhood information from historical censuses with adult outcome data from military records

Constructing the Dataset

- Individuals are sampled from World War II enlistment records
- These individuals are then matched to the 1930 federal census on the basis of name, birth year, and birth state
- If a match is found in the 1930 federal census, information on the individual's siblings and parents is recorded
- Resulting variables for successfully linked individuals include:
  - Enlistment records: height, weight, educational attainment, civilian occupation,...
  - Census records: siblings ages and genders, parents' occupations, house value, household location,...

Matching Enlistees to the Census

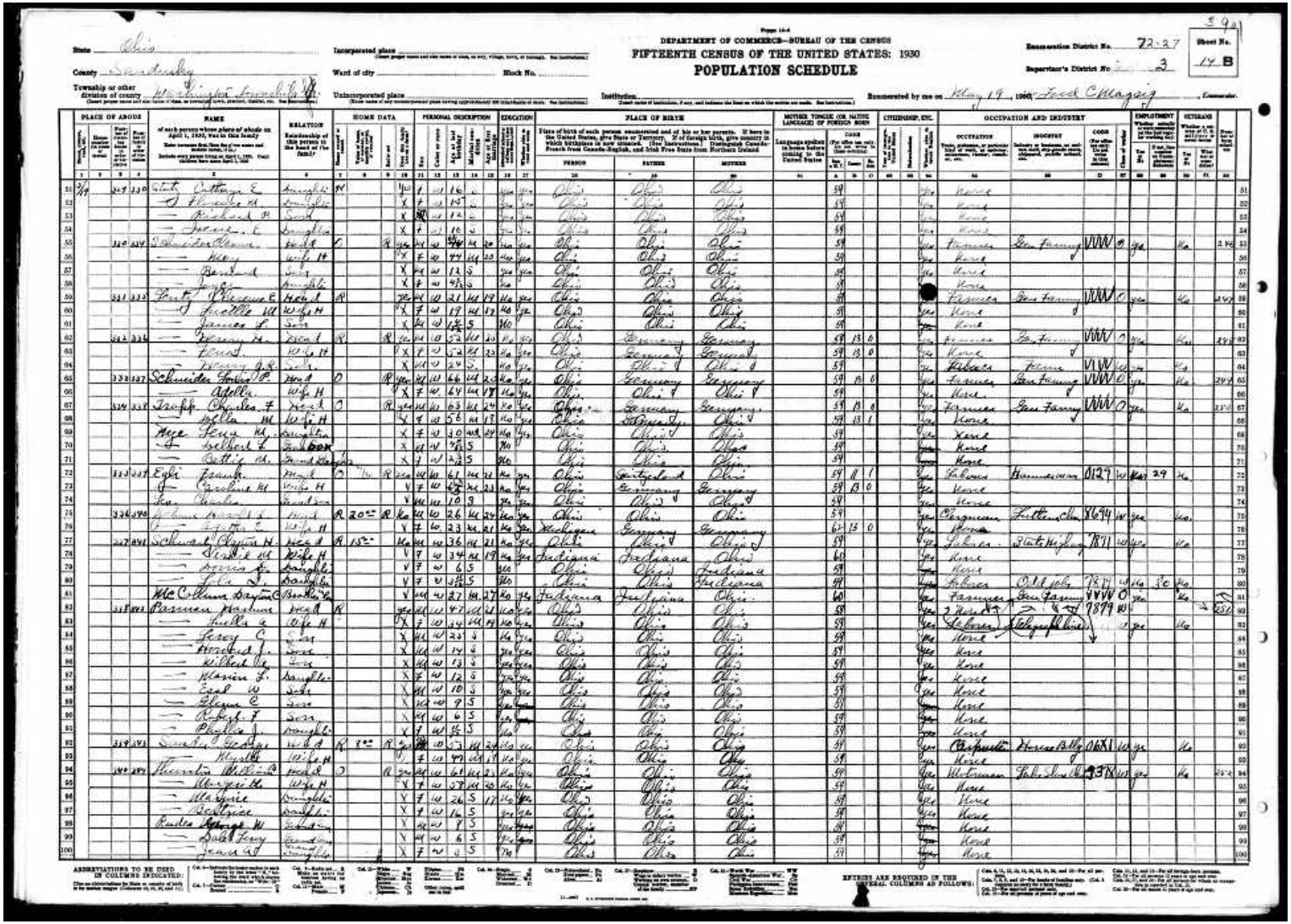
Field Title	Value	Meaning
ARMY SERIAL NUMBER	35548336	35548336
NAME	PARMAN#ROBERT#F##	PARMAN#ROBERT#F##
RESIDENCE: STATE	53	OHIO
RESIDENCE: COUNTY	123	OTTAWA
PLACE OF ENLISTMENT	5384	TOLEDO OHIO
DATE OF ENLISTMENT DAY	23	23
DATE OF ENLISTMENT MONTH	2	2
DATE OF ENLISTMENT YEAR	43	43
GRADE: ALPHA DESIGNATION	PVT#	Private
GRADE: CODE	8	Private
BRANCH: ALPHA DESIGNATION	BI#	Branch Immaterial - Warrant Officers, USA
BRANCH: CODE	0	Branch Immaterial - Warrant Officers, USA
FIELD USE AS DESIRED	#	#
TERM OF ENLISTMENT	5	Enlistment for the duration of the War or other emergency, plus six months, subject to the discretion of the President or otherwise according to law
LONGEVITY	###	###
SOURCE OF ARMY PERSONNEL	0	Civil Life
NATIVITY	53	OHIO
YEAR OF BIRTH	24	24
RACE AND CITIZENSHIP	1	White, citizen
EDUCATION	4	4 years of high school
CIVILIAN OCCUPATION	677	Semiskilled filers, grinders, buffers, and polishers (metal)
MARITAL STATUS	6	Single, without dependents
COMPONENT OF THE ARMY	7	Selectees (Enlisted Men)
CARD NUMBER	#	#
BOX NUMBER	957	957
FILM REEL NUMBER	5.19#	5.19#

Matching Enlistees to the Census

Matches 1-7 of 7 Sorted By Relevance

View Record	Name	Parent or spouse names	Home in 1930 (City, County, State)	Birth Year	Birthplace	Relation to Head of House	View Image
View Record ★★★★☆	Robert F Parman	Hannon, Luella A	Washington, Sandusky, Ohio	abt 1924	Ohio	Son	
View Record ★★★☆☆	Robert Parman	Omer, Martha	Nashville, Davidson, Tennessee	abt 1923	Tennessee	Son	
View Record ★★★☆☆	Robert L Parman	James J, Amy K	Washington, Harrison, Missouri	abt 1923	Missouri	Son	
View Record ★★★☆☆	Robert Parson [Robert Parman]	B L, Jennie	Melrose, Grundy, Iowa	abt 1926	Iowa	Son	
View Record ★★★☆☆	Robert R Parman	Elmer, Alma	Allen, Worth, Missouri	abt 1927	Missouri	Son	
View Record ★★★☆☆	Robert Carman [Robert Parman]	Claude, Edna	Brighton, Adams, Colorado	abt 1929	Iowa	Son	
View Record ★★★☆☆	Robert R Parman	Porter, Addie L	New Hampton, Harrison, Missouri	abt 1930	Missouri	Son	

Obtaining Information for Matched Enlistees

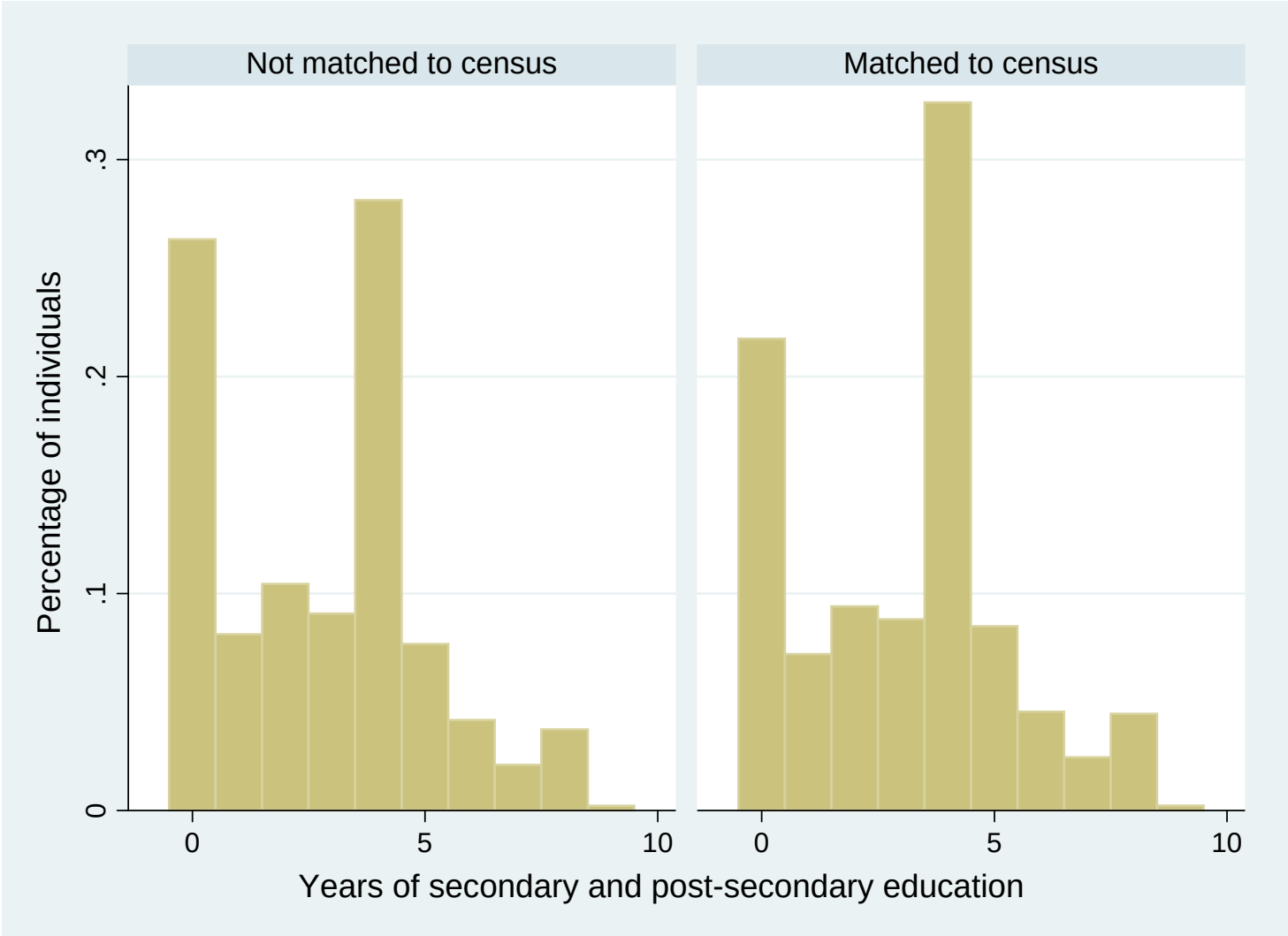


The Resulting Sample

Summary statistics by census match status

	Matched to federal census	Not matched to federal census
Number of individuals	13, 173	14,139
Year of birth	1918.8	1918.7
	(3.2)	(3.2)
Year of enlistment	1941.8	1941.8
	(0.8)	(0.9)
Height	68.4	68.3
	(2.7)	(2.8)
Weight	150.2	149.9
	(21.3)	(21.7)
Body mass index	22.5	22.5
	(2.9)	(3.0)
Years of secondary and post-secondary schooling	3.0	2.8
	(2.2)	(2.3)
Percentage who are white	94.4%	90.1%
Percentage who migrated to a different state	13.5%	25.0%
Percentage who are sons of the household head	90.1%	--

The Resulting Sample



Estimation Strategy

Given the adult outcomes from the enlistment records and the childhood household data from the census, estimation of the effect of the pandemic is fairly straightforward:

$$Y_{i,j} = f(X_i, Z_j, Flu_j, \varepsilon_{i,j})$$

- $Y_{i,j}$: adult outcome of interest (educational attainment, height, weight, BMI)
- X_i: individual characteristics (age, race, birth order, birth order among brothers, state of birth)
- Z_j: household characteristics (family size, parents' income, spacing and gender of siblings, house value, state of residence)
- Flu_j: indicator for sibling in utero during influenza pandemic

Estimation Strategy

The functional form will depend on the outcome of interest:

$$Y_{i,j} = f(X_i, Z_j, Flu_j, \varepsilon_{i,j})$$

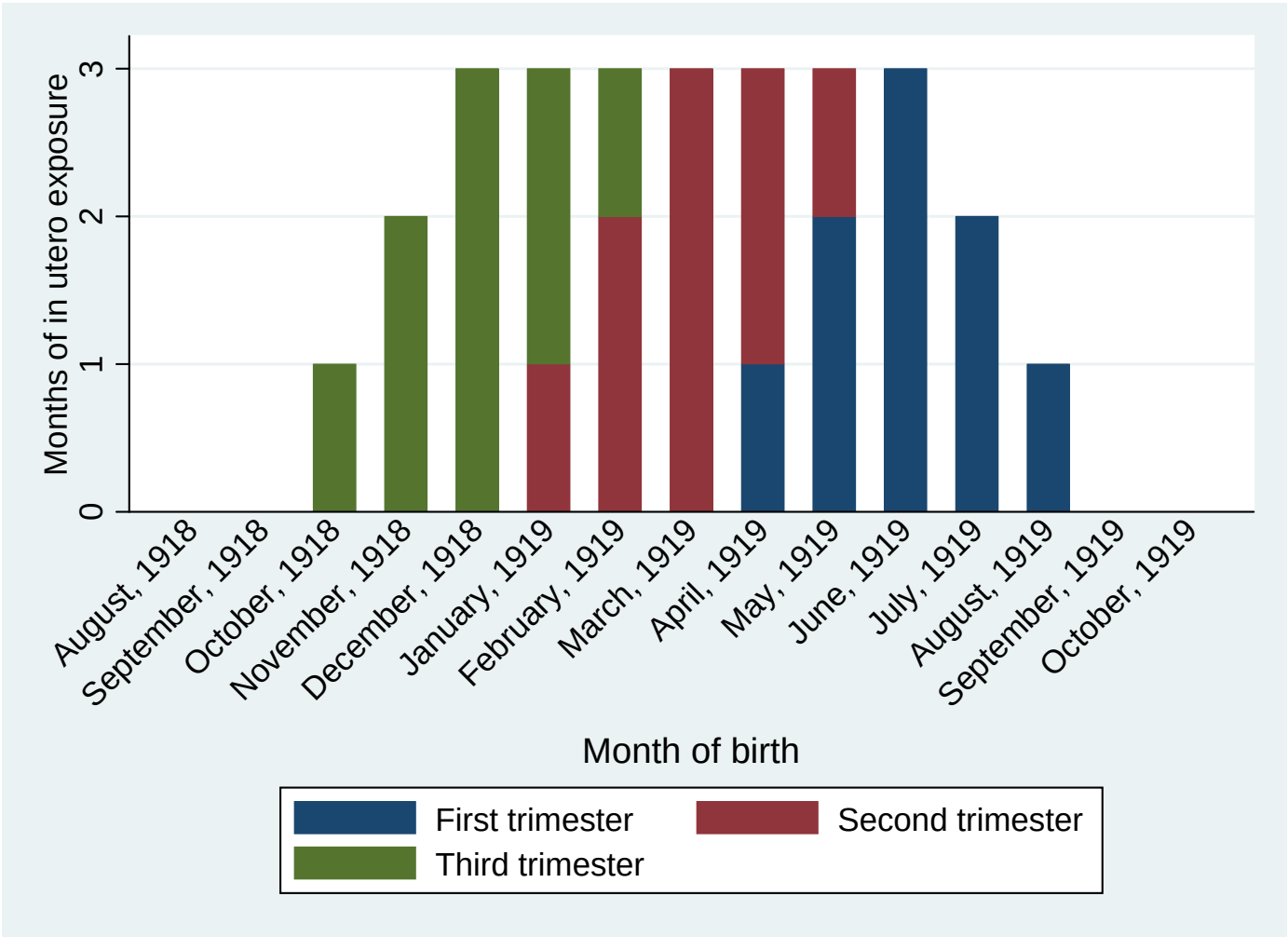
- Tobit model for years of education (over 20 percent of enlistees have zero years of secondary schooling)
- Logit models for binary education variables (attended HS, graduated HS)
- Linear models for health outcomes (height, weight, BMI)

Identifying Treated Siblings

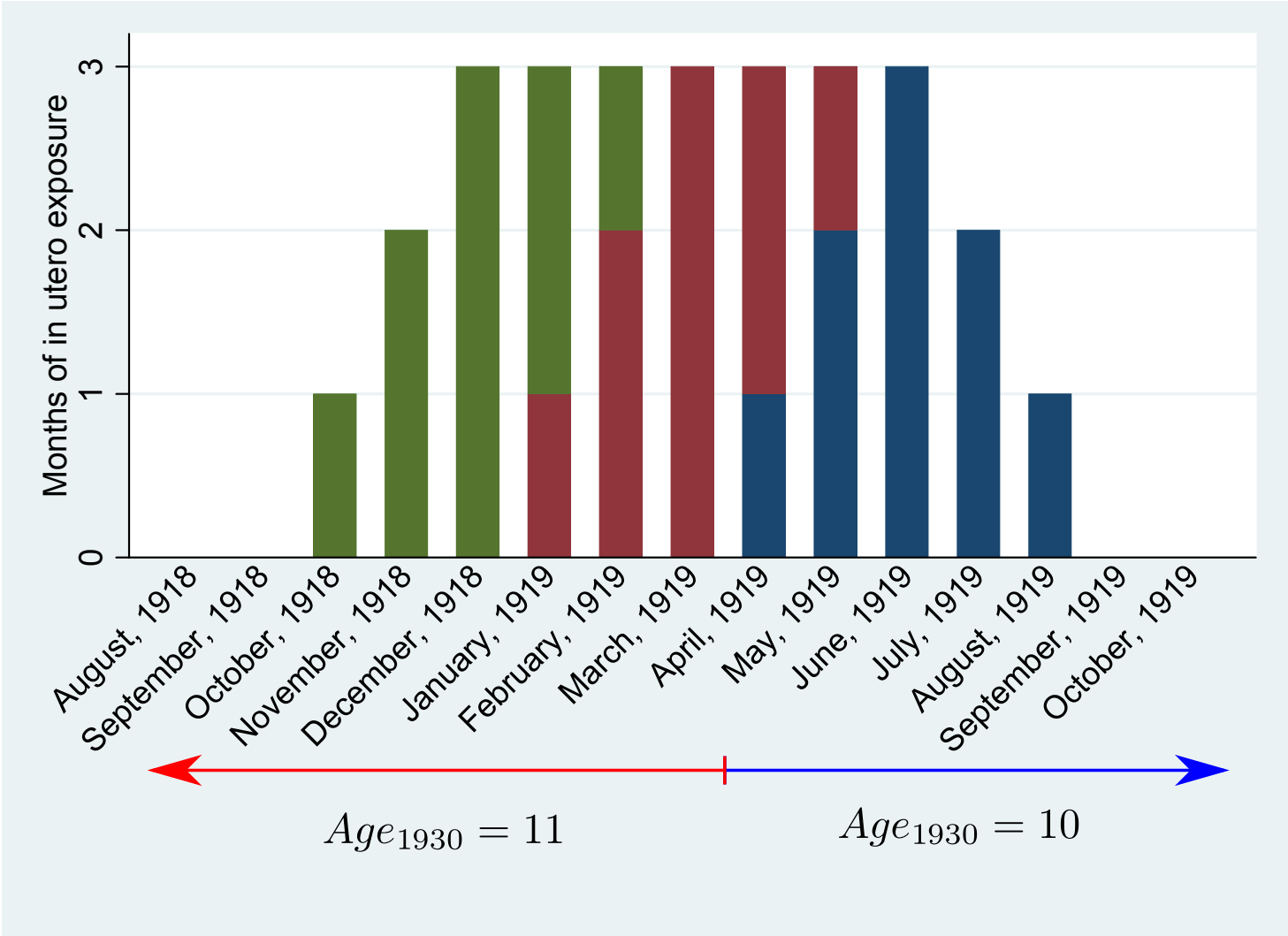
- The estimation strategy is based on identifying siblings in utero during the pandemic
- The severe months of the pandemic were October, November and December of 1918
- The birth cohorts in utero during the pandemic were therefore born in late 1918 and the first half of 1919
- Complicating things is the way age is reported in the federal census
- From the enumerator instructions:

  [Age at last birthday] calls for the age in completed years at last birthday. Remember, however, that the age question, like all other questions on the schedule, relates to April 1, 1930.

Identifying Treated Siblings



Identifying Treated Siblings



Summary Statistics for Regression Sample

Family characteristics of enlistees in regression sample

Variable	Mean	Standard deviation
Number of people in household	6.4	2.3
Number of siblings	3.0	2.2
Number of brothers	1.6	1.5
Number of older siblings	1.5	1.6
Number of older brothers	0.8	1.1
Percentage with an older sibling born in 1919	8.2%	
Percentage with a younger sibling born in 1919	6.9%	
Percentage with an older brother born in 1919	4.1%	
Percentage with a younger brother born in 1919	3.4%	

Outcomes of the Influenza Birth Cohort

Departure of 1919 male birth cohort outcomes from 1912-22 trend

	Results from Almond (2006)			Results using
Outcome	1960 Census	1970 Census	1980 Census	enlistee records
High school graduate	-0.021***	-0.020***	-0.014***	-0.028***
	(0.005)	(0.003)	(0.003)	(0.001)
Years of education	-0.150***	-0.176***	-0.117***	-0.132***
	(0.038)	(0.023)	(0.019)	(0.002)
Never attended high school	--	--	--	0.017***
	--	--	--	(0.001)
Height	--	--	--	-0.035***
	--	--	--	(0.007)
Weight	--	--	--	-0.371***
	--	--	--	(0.040)
Body Mass Index	--	--	--	-0.034***
	--	--	--	(0.006)
Observations	114,031	308,785	471,803	2,744,642

Results are the coefficient on an indicator for being in the influenza cohort included a linear regression controlling for birth year and birth year squared. Robust standard errors are given in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%

Effects of Sibling Health on Educational Attainment

Tobit regression results, years of secondary and post-secondary education as dependent variable

Sibling measure:	All siblings		Brothers only	
Includes individuals born in 1919:	no	yes	no	yes
Has older sibling born in 1919 (1=yes)	-0.038	-0.072	-0.145	-0.167
	(0.116)	(0.117)	(0.153)	(0.148)
Has younger sibling born in 1919 (1=yes)	0.267**	0.244**	0.244*	0.227*
	(0.113)	(0.109)	(0.142)	(0.134)
Number of older siblings	-0.172***	-0.169***	-0.196***	-0.196***
	(0.021)	(0.019)	(0.031)	(0.029)
Number of younger siblings	-0.359***	-0.352***	-0.426***	-0.422***
	(0.020)	(0.017)	(0.034)	(0.030)
Household head's log income	1.665***	1.627***	1.756***	1.719***
	(0.117)	(0.111)	(0.119)	(0.113)
Observations	8,743	9,865	8,743	9,865

Robust standard errors clustered by birth state in parentheses. Regressions include controls for race, state fixed effects, presence of the father and a cubic in age. * significant at 10%; ** significant at 5%; *** significant at 1%

Effects of Sibling Health on Educational Attainment

Logit regression results, high school attendance or graduation as dependent variable

Dependent variable:	Attended high school		Graduated high school	
Includes individuals born in 1919:	no	yes	no	yes
Has older sibling born in 1919 (1=yes)	-0.026	-0.053	-0.122	-0.148
	(0.100)	(0.102)	(0.105)	(0.105)
Has younger sibling born in 1919 (1=yes)	0.355***	0.333***	0.233***	0.223***
	(0.104)	(0.101)	(0.085)	(0.084)
Number of older siblings	-0.124***	-0.123***	-0.131***	-0.123***
	(0.018)	(0.017)	(0.021)	(0.018)
Number of younger siblings	-0.240***	-0.234***	-0.265***	-0.263***
	(0.016)	(0.015)	(0.019)	(0.017)
Household head's log income	1.393***	1.372***	1.007***	0.990***
	(0.079)	(0.078)	(0.086)	(0.082)
Observations	8,743	9,865	8,743	9,865

Robust standard errors clustered by birth state in parentheses. Regressions include controls for race, state fixed effects, presence of the father and a cubic in age. * significant at 10%; ** significant at 5%; *** significant at 1%

Effects of Sibling Health on Educational Attainment

Marginal effects (dP(y=1)/dx) of siblings on educational outcomes based on logit regressions

Dependent variable:	Attended high school	Graduated high school
Mean of dependent variable:	0.792	0.524
Has older sibling born in 1919	-0.005	-0.030
	(0.018)	(0.026)
Has younger sibling born in 1919	0.056	0.058
	(0.015)	(0.021)
Number of older siblings	-0.021	-0.032
	(0.003)	(0.005)
Number of younger siblings	-0.042	-0.065
	(0.003)	(0.005)
Household head's log income	0.241	0.249
	(0.013)	(0.021)

Effects of Sibling Health on Educational Attainment

OLS estimates of sibling health on own adult health outcomes

Dependent variable:	Height	Weight	Body mass index
Includes individuals born in 1919:	no	no	no
Has older sibling born in 1919 (1=yes)	0.055	0.497	0.068
	(0.110)	(0.568)	(0.096)
Has younger sibling born in 1919 (1=yes)	0.081	0.182	-0.026
	(0.089)	(0.613)	(0.093)
Number of older siblings	-0.073***	-0.473***	-0.021
	(0.028)	(0.138)	(0.017)
Number of younger siblings	-0.129***	-0.756***	-0.019
	(0.025)	(0.103)	(0.018)
Household head's log income	0.192*	0.983	0.021
	(0.103)	(0.832)	(0.113)
Observations	8,743	8,624	8,624

Robust standard errors clustered by birth state in parentheses. Regressions include controls for race, state fixed effects, presence of the father and a cubic in age. * significant at 10%; ** significant at 5%; *** significant at 1%

Falsification Tests

- The results suggest strong effects of having a younger sibling born during the pandemic
- However, there is the possiblity that this sibling variable is a proxy for something else related to family structure and the enlistee's birth year
- Two approaches will help alleviate concerns:
  - Rerun everything with the inclusion of birth cohort dummies to more flexibly control for cohort trends
  - Rerun analysis using placebo influenza cohorts by assuming influenza pandemic was earlier or later

Falsification Tests

Tobit sibling coefficients for placebo influenza cohorts

Influenza birth year defined as:	1919	1917	1921
Includes individuals born in flu cohort:	no	no	no
Has older sibling born in flu cohort (1=yes)	-0.038	0.001	-0.139
	(0.116)	(0.102)	(0.174)
Has younger sibling born in flu cohort (1=yes)	0.267**	-0.061	-0.035
	(0.113)	(0.156)	(0.087)
Number of older siblings	-0.172***	-0.173***	-0.171***
	(0.021)	(0.017)	(0.020)
Number of younger siblings	-0.359***	-0.345***	-0.344***
	(0.020)	(0.021)	(0.018)
Household head's log income	1.665***	1.617***	1.626***
	(0.117)	(0.116)	(0.111)

Robust standard errors clustered by birth state in parentheses. Regressions include controls for race, state fixed effects, presence of father and a cubic in age. * significant at 10%; ** significant at 5%; *** significant at 1%

Falsification Tests

Logit sibling coefficients for placebo influenza cohorts (attended high school as dependent variable)

Influenza birth year defined as:	1919	1917	1921
Includes individuals born in flu cohort:	no	no	no
Has older sibling born in flu cohort (1=yes)	-0.026	0.008	-0.224
	(0.100)	(0.093)	(0.170)
Has younger sibling born in flu cohort (1=yes)	0.355***	-0.022	-0.043
	(0.104)	(0.119)	(0.078)
Number of older siblings	-0.124***	-0.127***	-0.120***
	(0.018)	(0.016)	(0.019)
Number of younger siblings	-0.240***	-0.223***	-0.216***
	(0.016)	(0.016)	(0.014)
Household head's log income	1.393***	1.380***	1.344***
	(0.079)	(0.083)	(0.088)

Robust standard errors clustered by birth state in parentheses. Regressions include controls for race, state fixed effects, presence of father and a cubic in age. * significant at 10%; ** significant at 5%; *** significant at 1%

Falsification Tests

Logit sibling coefficients for placebo influenza cohorts (graduated high school as dependent variable)

Influenza birth year defined as:	1919	1917	1921
Includes individuals born in flu cohort:	no	no	no
Has older sibling born in flu cohort (1=yes)	-0.122	-0.030	-0.161
	(0.105)	(0.084)	(0.114)
Has younger sibling born in flu cohort (1=yes)	0.233***	-0.106	-0.041
	(0.085)	(0.096)	(0.062)
Number of older siblings	-0.131***	-0.128***	-0.116***
	(0.021)	(0.016)	(0.017)
Number of younger siblings	-0.265***	-0.258***	-0.252***
	(0.019)	(0.017)	(0.017)
Household head's log income	1.007***	0.978***	0.979***
	(0.086)	(0.087)	(0.081)

Robust standard errors clustered by birth state in parentheses. Regressions include controls for race, state fixed effects, presence of father and a cubic in age. * significant at 10%; ** significant at 5%; *** significant at 1%

Conclusions, Future Work

- The sibling data reveal significant reallocation of resources within families with a major health shock
- Results also suggest that effects on older siblings are quite different than effects on subsequent siblings
- These impacts are big: the gain in education from having a sick sibling is similar in magnitude to the gain associated with a 20 percent increase in parental income
- Several future directions for the work:
  - Household resource allocation models that can account for observed patterns
  - Implications for intergenerational mobility
  - Make use of the 1940 federal census (income, occupations, educational attainment for all siblings; ability to include females in regressions)