Health effects of cousin marriage: Evidence from US genealogical records*

Sam Il Myoung Hwang[†], Deaglan Jakob[†], Munir Squires[†]

April 8, 2024

Abstract

Cousin marriage rates are high in many countries today. We provide the first estimate of the effect of such marriages on the life expectancy of offspring. By studying couples married over a century ago, we can observe their offspring across the lifespan. US genealogical data allows us to identify children whose parents were first cousins, and compare their years of life to the offspring of their parents' siblings. Marrying a cousin leads to more than a two-year reduction in age-5 life expectancy. This effect is strikingly stable across time, despite large changes in life expectancy and economic environment.

*This paper has greatly benefited from the advice and suggestions of Siwan Anderson, Nathan Nunn, and Marit Rehavi. We also thank Jeff Chan, Karen Clay, Mauricio Drelichman, Shari Eli, Claudio Ferraz, Patrick Francois, Matt Lowe, Laura Salisbury, as well as CEA, CNEH, and VSE seminar participants for numerous useful comments. This research was undertaken, in part, thanks to funding from the Canada Excellence Research Chairs program awarded to Dr. Erik Snowberg in Data-Intensive Methods in Economics.

[†]University of British Columbia, Vancouver School of Economics. Hwang: hwangii@mail.ubc.ca. Jakob: deaglanj@mail.ubc.ca. Squires: munir.squires@ubc.ca

1 Introduction

The taboo against cousin marriage in the US and Europe is linked to the belief that children from these unions are likely to have genetic health problems. In many societies today, however, marriages between first or second cousins are common. A much-cited estimate is that these couples and their offspring make up 10% of the world's population (Bittles and Black, 2010). In some countries the rate is much higher: about 50% of marriages in Pakistan are between first cousins.[1]

Biologists and medical researchers have converged on the conclusion that the effect of cousin marriage on offspring health is real but modest in magnitude (Bittles and Neel, 1994; Bittles and Black, 2010). According to an influential report by the US National Society of Genetic Counsellors, "There is a great deal of stigma associated with cousin unions in the United States and Canada that has little biological basis" (Bennett et al., 2002). The report concludes that the risks are smaller than assumed and do not justify additional genetic testing.[2]

This paper documents large health effects from first-cousin marriage. This is largely because we observe mortality beyond childhood, throughout the lifespan. Doing so requires studying couples who were married a hundred or more years ago as well as their offspring. Since this type of data is not available in countries with high rates of cousin marriage today, we turn to historical US genealogical records to fill this gap. Our data allows us to directly identify first cousin marriages and measure years of life lived by the offspring of these and other marriages.

We first show a strikingly robust pattern of lower life expectancy for offspring of first-cousin marriages. This gap is consistent across birth cohorts from 1750-1900, and throughout the distribution of parental longevity.

To determine whether this difference reflects the causal effect of cousin marriage, we compare the children of married cousins to the children of the *siblings* of these married cousins. This empirical approach has the advantage of controlling for a wide range of potential genetic, economic and cultural sources of selection into cousin marriage. While

[1] The share of ever-married women ages 15-49 who report marrying a first cousin in the Pakistan DHS survey was 49.6% in the most recent 2017-18 round. Younger women report higher rates of cousin marriage, suggesting the practice is not in decline.

[2] This report was widely cited in debates on cousin marriage sparked by members of the UK parliament (Paul and Spencer, 2008). In an article discussing updated guidelines (Bennett et al., 2021), the lead author of this report is quoted as saying that "the risks are very low and not much different than for any other couple." (The Economist, 2024)

this adjusts for unobserved characteristics shared by sets of siblings, differences between siblings who marry a cousin and those who do not may still lead to selection bias. If siblings in worse health are more likely to marry a cousin, their children may live shorter lives for reasons unrelated to consanguinity. We test this by comparing the longevity of individuals who marry a cousin to their siblings. Reassuringly, we find no difference: within sets of siblings, marrying a cousin is not correlated with one's own longevity.

Our main result is that life expectancy at age 5 is more than two years shorter for children with first-cousin parents. Notably, we also find that mortality effects accrue throughout the lifespan.

While our genealogical dataset includes millions of individuals, it is not representative of populations for whom relatively few historical records exist. Notably, immigrants and non-whites are almost completely excluded from our data. Another limitation of our genealogical data is that about two thirds of infant deaths are missing. Because of this our main analysis focuses on life expectancy at age 5. Since child mortality rates are higher for married cousins, this leads us to substantially underestimate the associated decline in life expectancy. Accounting for child mortality suggests that cousin marriage reduces life expectancy at birth by at least three years.

The most straightforward channel through which cousin marriage reduces offspring lifespan is genetic. However, our data do not allow us to directly test for genetic versus other mechanisms. Notably, in previous work we show that cousin marriage reduces geographic and occupational mobility, which results in lower incomes (Ghosh, Hwang and Squires, 2023). This could itself reduce the health of offspring.

This paper adds to a literature in economics on the determinants of health, which documents the importance of health stock on economic outcomes (see for example Currie et al., 2009; Strauss and Thomas, 2007, and citations therein). We contribute to this literature by documenting the health costs of a practice that, while now rare in developed countries, is widely practiced in many societies.

We also advance the multi-disciplinary literature on the health effects of cousin marriage (see Hamamy et al., 2011; Bittles and Black, 2010, for reviews of this literature). Our main contribution is to provide the first estimate of its effects on life expectancy. We can do this because we study individuals born sufficiently long ago that we can follow them throughout their lifespan.[3] Second, we address selection into cousin marriage by restricting comparisons

[3] Existing studies focus on infant or child mortality, which is more easily measured. There is significantly more disagreement on the effect of cousin marriage on adult health outcomes. While some studies find a

to close relatives who share genetic and other unobserved characteristics.[4] The only other paper in this literature to address selection on unobservables is Mobarak et al. (2019), who use variation in the availability of marriageable opposite-sex cousins as an instrument. They find modest but noisy increases in child mortality.

A related study, Ghosh et al. (2023), explores the economic consequences of cousin marriage. Using variation from US state bans on cousin marriage, it finds that such marriages reduce geographic and occupational mobility. We contribute to our understanding of the effects of cousin marriage by studying health, rather than economic, outcomes using a new data source (genealogical profiles).

This paper also contributes to an economics literature on the role of culture in shaping marriage and family decisions (Fernández, 2011; Giuliano and Nunn, 2021). While there is evidence that cousin marriage plays a functional role,[5] anthropologists have long emphasized that cousin marriage is a deeply cultural practice (Lévi-Strauss, 1969). In that vein, our results add to the literature on the health and economic costs of cultural practices (Lowes and Montero, 2021; Corno et al., 2020; Atkin, 2016; Almond and Mazumder, 2011).

2 Genealogical data

Our measures of longevity and family ties come from FamilySearch, a genealogical website where users can view historical records and enter information about their ancestors. A main feature of the site is a set of public genealogical profiles of historical individuals, creating large, interlinked family trees.[6] The dataset we use in this paper consists of 40 million of these linked individual profiles. Using these, we trace the genealogies of individuals to identify cousin marriages and study their effect on longevity.

Our sample was obtained by first collecting all US marriage records up to the mid-nineteenth century available on FamilySearch. Using the profiles of these spouses, we

negative effect (Rudan et al., 2003a,b; Bener et al., 2007; Liede et al., 2002; Gilani et al., 2006), others find none (Bener et al., 2009; McWhirter et al., 2012; Bener et al., 2010; Denic et al., 2007). Like us, Helgason et al. (2008) and Kaplanis et al. (2018) use large-scale genealogical data to identify cousin marriage, though neither studies its effects on life expectancy.

[4]Our approach is similar to a sibling comparison, commonly used to address selection challenges (Abramitzky et al., 2012; Collins and Wanamaker, 2014; Ward, 2022; Lu and Vogl, 2023; Kreisman and Smith, 2023). One distinction is that we compare the *offspring* of siblings, rather than the siblings themselves.

[5]For example by managing inheritance (Bahrami-Rad, 2021) or by making commitments between the groom and bride's families more credible (Mobarak et al., 2013).

[6]See Price et al. (2021); Hwang and Squires (2024a); Blanc (2023) for evidence on the quality of this type of genealogical data. Appendix Figure A.1 shows a sample genealogical profile.

expanded our sample both horizontally (siblings and siblings-in-law), and vertically (parents and children). Our resulting sample is about half the size of the early-mid-nineteenth-century US population (Appendix Figure A.2).

This section describes two key features of this dataset. First, for about a quarter of our sample we have sufficient vertical genealogical links to determine whether their parents were first cousins. Second, these profiles include birth and death years to measure longevity, which we use as a summary measure of health. However, our data come with potential challenges: infant deaths are underreported, vital dates may be recorded with error, and our sample underrepresents non-whites and immigrants. We address these below.

To identify a cousin marriage, we need to observe all four grandparents of both spouses. For the couples in our data for whom we have all eight grandparents, we determine whether spouses are cousins by checking for overlap in the two sets of grandparents. The vast majority of spouses in our data have either zero (97.3%) or two (2.5%) overlapping grandparents.[7] Spouses with two grandparents in common are first cousins, while those with no matching grandparents are not (first) cousins. Appendix Figure A.3 illustrates the family tree of two spouses who are first cousins. Appendix Figure A.4 shows that first-cousin marriage measured in this way declines in our data from about three to one percent of marriages from 1750 to 1900.

As a robustness check, we also provide results using an alternative measure of cousin marriage: same-surname spouses, defined as marriages where the husband's surname matches that of the wife's father. This method, although less precise than our primary measure, allows for a broader sample as it does not require information on all eight grandparents of a given couple. As discussed in Ghosh et al. (2023), same-surname spouses can be a meaningful proxy of cousin marriage, as first cousins are far more likely to share a surname than two unrelated spouses.

To evaluate how marrying a cousin affects the health of offspring, we use as an outcome their years of life ('longevity'). Since genealogical profiles do not contain direct information on health, such as diseases, disability, or cause of death, we treat longevity as a proxy for overall, life-time health. Measuring longevity simply requires us to take the difference between individuals' birth and death years, which are available for about three-fourths of

[7]We omit the 0.2% of couples with one, three or four matching grandparents from our analysis for simplicity and because of insufficient statistical power. Having one matching grandparent would mean the spouses are half-first-cousins, and having three or four implies they are half or full siblings, or double first cousins.

our sample.

A major challenge with using genealogical records to estimate life expectancy is that offspring who die young are often missing. Infant mortality in our data is about two-thirds smaller than benchmark figures from Hacker (2010).[8] Because of this, our baseline results condition on children having reached age 5. That is, we report the effect of cousin marriage on life expectancy at age 5 (and 20) rather than at birth. In supplementary analysis we confirm that conditioning on survival to age 5 means we substantially understate the overall effect on life expectancy.

The sample we use for analysis meets two additional criteria: non-missing data on all eight great-grandparents (their mother's and father's four grandparents), and non-missing birth and death years. We also drop individuals with missing data or errors in their genealogical profiles, and restrict our sample to those born between 1750 and 1920.[9] Of the 40 million individuals in our dataset, the 6 million who meet these criteria form our 'analysis sample' (Appendix Table A.1).

A second weakness of our dataset is that vital years may be recorded with error. Indeed, we observe heaping in death years, which unlike birth years cannot be obtained from census records. This heaping is likely because, in the absence of alternative records, genealogical researchers use the last census year where an individual was observed as their year of death. We show that these errors do not bias our estimates.

A final weakness of our data is that, while broad in scope, the 40 million profiles we use are not representative of the US population. Crowd-sourced genealogical data typically under-represent people for whom fewer records exist (Price et al., 2021; Hwang and Squires, 2024a; Buckles et al., 2023). Our analysis sample, which consists of the 6 million individuals for whom we have genealogical records for all eight great-grandparents, is likely to be even less representative.

Quantifying how similar our analysis sample is to the census population is complicated by the fact that the genealogical profiles we use do not include race, occupation or literacy. In Appendix Table A.3 we present suggestive evidence on how our sample differs from the

[8] See Appendix Figure A.5, where we compare age-specific mortality rates in our sample to this benchmark. To our knowledge, Hacker (2010) provides the best available estimates of mortality rates during our period of interest, though this is based on extrapolations of non-representative sub-samples.

[9] Dropping individuals born after 1920 limits potential selection out of sample by individuals who are still alive. We drop individuals with missing sex or maternal age at birth, and those whose longevity is negative or above the 99th percentile (98 years old). Individuals with impossible familial links (e.g., children being their own parents) are dropped from the sample. To provide a consistent sample across analyses, we also drop the 0.1 million singleton observations that get dropped from our fixed effects estimation.

census population using a subsample of profiles that we can link to the 1850 census. We find that our analysis sample consists almost exclusively of individuals who are white, literate and US-born. Farmers are over-represented, as are those born in the Northeast. We think of our sample, then, as broadly representative of the white, US-born population.

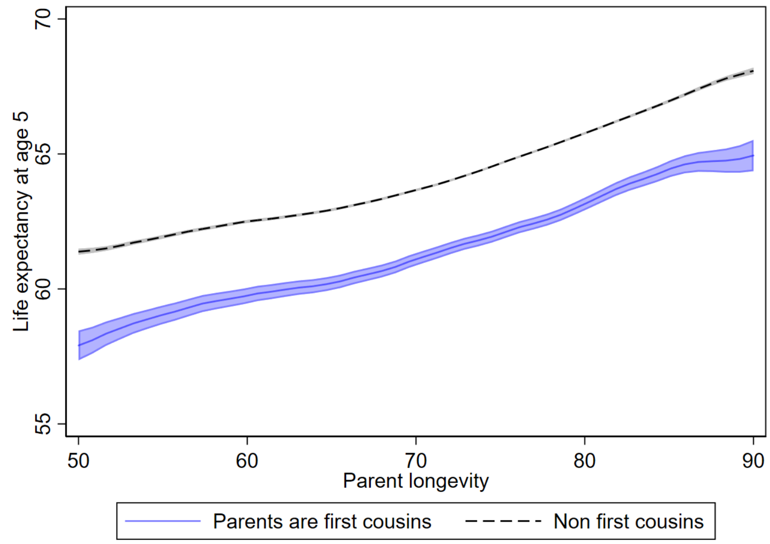
Descriptive statistics of our analysis sample are presented in Appendix Table A.2. We split the sample between children of married cousins (2.5% of the sample) and non-cousins (97.5%) to illustrate basic differences between these two groups. Most notably, conditional on reaching age 5 the average longevity of offspring of cousin marriages is 61 years, relative to 64 years for the offspring of non-cousins, a three-year difference. Parental longevity, defined as the average of the mother and father's longevity, is also shorter. Married cousins live half a year shorter lives. If individuals who marry their cousins are themselves in worse health, then the health of their offspring cannot easily be compared to that of the population at large.

Before presenting our empirical design, which aims to address potential selection into cousin marriage, we show in Figure 1 that this longevity gap is strikingly stable. Panel (a) shows how offspring life expectancy changes with parental longevity. As expected, the offspring of longer-lived parents also live longer (Black et al., 2023). Differences in parental longevity presumably reflect a combination of genetic, socio-economic, and geographic differences between families. The figure highlights that the difference in life expectancy for offspring of married cousins is large and stable across the parental longevity distribution.

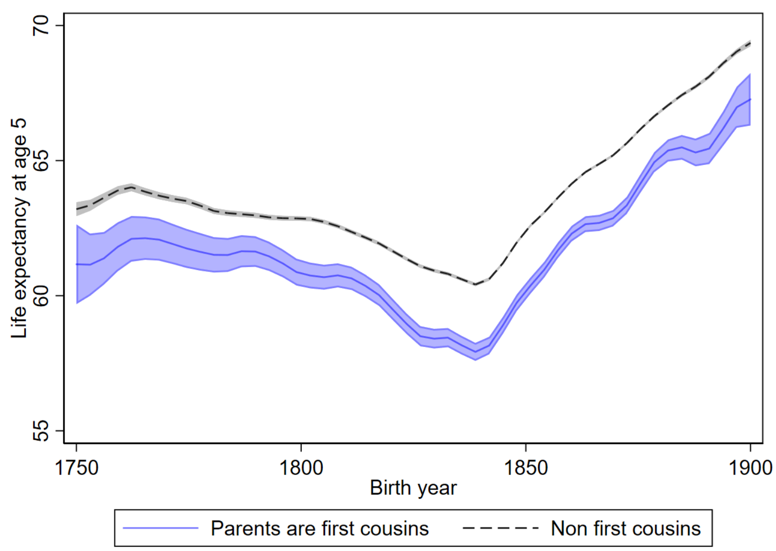
Panel (b) of Figure 1 likewise documents that this difference is stable across birth cohorts. While cohort life expectancy changed substantially in our dataset between 1750 and 1900, the gap for offspring of first-cousin parents remains stable.[10] Offspring of cousin marriages have consistently lower life expectancy than those born of non-cousins, despite dramatic economic and social changes in the US during this period. We also find that this difference is stable across locations (Appendix Figures A.7 and A.8) despite large differences in mortality (Finkelstein et al., 2021).

[10] The broad pattern of life expectancy at birth in our data is consistent with Hacker (2010). Notice that the US Civil War had a large negative effect on life expectancy, especially for individuals born in the 1830s and 40s. The gap in adult life expectancy is also large and stable across birth cohorts (Appendix Figure A.6).

Figure 1: Cousin marriage and life expectancy



(a) Life expectancy at age 5 by parental longevity



(b) Life expectancy at age 5 by birth cohort

This figure depicts the life expectancy at age 5 of our analysis sample of 5.9 million offspring. In both panels, offspring of first cousins are represented by solid lines and offspring of non first cousins are represented by dashed lines. Panel (a) is a local polynomial regression of life expectancy on parent longevity. Parent longevity is the average of the mother's and father's longevity. Panel (b) is a local polynomial regression of life expectancy on birth year. The shaded areas represent 95% confidence intervals.

3 Empirical design

3.1 Regression specification

Our analysis studies children of married cousins, and compares them to the children of their parents' siblings. We estimate the effect of cousin marriage on years of life (*Longevity*) of children born of these marriages using the following empirical specification:

$$Longevity_i = \beta FirstCousinParents_i + \lambda_{m(i)} Maternal_{m(i)} + \lambda_{p(i)} Paternal_{p(i)} + \mathbf{X}_i^{\prime}\boldsymbol{\delta} + \epsilon_i, \quad (1)$$

where i is an individual in our analysis sample. The treatment variable $FirstCousinParents_i$ is equal to 1 if i's parents are first cousins, and 0 if not. Subscripts m and p denote maternal (mother's side) and paternal (father's side) relatives of i. Specifically, each individual i shares a value of m with all children of i's mother and maternal aunts (mother's sisters). Likewise, i shares a value of p with all children of i's father and his brothers. $Maternal_{m(i)}$ is equal to one for all individuals with the same value of m, and zero otherwise, as with $Paternal_{p(i)}$. Appendix Figure A.9 illustrates the relevant comparison groups for i implied by the $Maternal_{m(i)}$ and $Paternal_{p(i)}$ fixed effects. Finally, $\mathbf{W}_j$ is a vector of individual-level controls that consists of an indicator for being female, as well as quadratic controls for birth year, maternal age at birth, number of sisters, number of brothers, sibling sex ratio, and birth order.[11] Standard errors are clustered at the level of siblings.

The maternal and paternal fixed effects allow us to restrict comparisons to close relatives. These serve as a useful control group as they share a wide range of unobserved economic, social, and genetic characteristics. Within-family comparisons address potential concerns that arise out of non-random selection of families (sets of siblings) into higher or lower rates of cousin marriage.[12] Our key identifying assumption is the following: *within* sets of siblings, selection into cousin marriage is independent of other traits that might affect offspring longevity. The following section tests this assumption.

[11] Maternal age at birth is truncated below 12 and above 50. Number of sisters (brothers) includes all female (male) offspring of one's parents, including oneself. These two variables and birth order are top-coded at 10. Sex ratio is the number of sisters divided by the total number of siblings.

[12] This approach addresses another potential concern: marrying a cousin might also mean choosing a spouse with traits linked to higher or lower offspring longevity. The inclusion of maternal *and* paternal fixed effects means the choice of spouse affects who one's children will be compared to.

3.2 Test of key identifying assumption

Our data allow us to directly test whether the longevity of individuals who marry a cousin differs from their siblings. To do so, we implement an empirical specification similar in principle to equation (1), but where the units of observation are parents of those in the analysis sample, and the outcome is the longevity of these parents. Appendix Table A.4 presents descriptive statistics on the parent-level observations we use in this section.

Namely, we use the following specification:

$$Longevity_j^{Parental} = \gamma FirstCousinSpouse_j + \eta_{s(j)} SameSexSiblings_{s(j)} + \mathbf{W}'_j \boldsymbol{\phi} + \varepsilon_j. \quad (2)$$

Here, j refers to a parent of one of the individuals in our main analysis sample. Subscript s refers to a set of same-sex siblings (a father and his brothers, a mother and her sisters). *FirstCousinSpouse* is equal to 1 if that parent married a first cousin, and 0 otherwise. Parents who had children with more than one spouse are treated as having married a cousin if any one of those spouses was a cousin. The set of controls $\mathbf{W}_j$ consists of an indicator for female, as well as quadratic controls for birth year, maternal age at birth, number of sisters, number of brothers, sibling sex ratio, and sibling birth order.

Column (1) of Table 1 reports that parents who married their cousin live 0.6 fewer years. Individual-level controls in column (2) reduce the estimated coefficient slightly.

This difference disappears entirely when we include sibling fixed effects in column (3) of Table 1. Adding these fixed effects means that we compare the longevity of individuals to their same-sex siblings. That is, fathers are compared to their brothers, and mothers to their sisters. Restricting to within-sibling comparisons reduces the coefficient to zero, with a reasonably precise confidence interval.[13]

Importantly, we are not claiming that the effects reported in Table 1 are causal: indeed we are interested in selection into cousin marriage, and wish to test whether siblings who marry their cousins are in worse health. That we find no correlation with longevity after including sibling fixed effects suggests that within-sibling comparisons adequately control for the relevant confounding variables. Restricting our comparisons to the children of these same-sex siblings should therefore allow us to recover the causal effect of cousin marriage on the health of offspring.

[13] Similarly, for a small sub-sample of siblings linked to the census, Appendix Table A.5 compares married cousins and non-cousins. With the addition of sibling fixed effects, none of the differences in place of birth, literacy, occupation and wealth are statistically significant.

Table 1: Selection of parents into cousin marriage

	(1) Raw	(2) Controls	(3) Same-sex sibling fixed effects
Parent Longevity			
Married to first cousin	-0.57	-0.45	-0.00
	(0.09)	(0.09)	(0.12)
Control mean	67.95	67.95	67.95
Observations (thousands)	1,255	1,255	1,255
Controls	No	Yes	Yes
Same-sex sibling FE	No	No	Yes

This table shows estimates for the coefficient γ from equation (2) estimated using OLS. Each observation is a parent of one of the offspring in our analysis sample. The outcome is that parent's longevity (year of death minus year of birth). Column (1) coefficients are simply the difference in means between those who marry their first cousins and those who do not. Column (2) adds controls for birth year, sex, maternal age at birth, number of sisters, number of brothers, the sex ratio of siblings, and birth order, as described in section 3. Column (3) adds same-sex sibling fixed effects (a father and his brothers, a mother and her sisters). Standard errors are clustered at the level of the individual and their siblings.

Table 2: The effect of cousin marriage on offspring longevity

	Baseline specifications			Robustness checks				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Raw	Controls	Parental fixed effects	Flexible controls	Parent longevity	Age heaping	County-decade FE	Same surname
Panel A: Life expectancy at age 5								
Parents are first cousins	-2.77	-2.19	-2.21	-2.20	-2.10	-2.09	-2.07	-1.86
	(0.07)	(0.07)	(0.29)	(0.29)	(0.30)	(0.32)	(0.31)	(0.19)
Control mean	63.73	63.73	63.73	63.73	63.73	63.86	64.14	63.70
Observations (thousands)	5,915	5,915	5,915	5,915	5,845	5,186	4,578	12,750
Panel B: Life expectancy at age 20								
Parents are first cousins	-2.40	-1.87	-1.81	-1.83	-1.64	-1.53	-1.95	-1.52
	(0.06)	(0.06)	(0.26)	(0.26)	(0.28)	(0.29)	(0.33)	(0.17)
Control mean	66.75	66.75	66.75	66.75	66.75	66.87	67.07	66.77
Observations (thousands)	5,569	5,569	5,569	5,569	5,504	4,895	4,316	12,034
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Paternal fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Maternal fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes

This table shows estimates for the coefficient β from equation (1) estimated using OLS. Each observation is an offspring in the analysis sample. The outcome is the child's longevity (year of death minus year of birth), conditional on surviving to a specified age. Column (1) coefficients are simply the difference in mean longevity between the children of first cousins and the children of non-first cousins. Column (2) adds controls for birth year, sex, maternal age at birth, number of sisters, number of brothers, the sex ratio of siblings, and birth order, as described in section 3. Column (3) adds mother's siblings fixed effects and father's siblings fixed effects. Column (4) replaces the quadratic controls with sets of fixed effects for each integer value. Column (5) controls for parent longevity. Column (6) drops all individuals with death dates ending in 0. Column (7) adds county-by-decade-of-birth fixed effects and removes controls for birth year. In column (8), the treatment variable is equal to 1 if the child's parents have an identical surname, and 0 otherwise. We use the mother's father's surname instead of her own surname to account for the fact that she may have taken her husband's name in marriage. Standard errors are clustered at the level of the individual and their siblings.

4 Results

What effect does cousin marriage have on the life expectancy of offspring? Table 2 presents results from OLS regressions that report the difference in longevity between children of cousins and of non-cousins. Each observation is a person ('offspring') in our analysis sample. Column (1) of Panel A reports that life expectancy at age 5 is on average 2.8 years lower for individuals born of first cousins. This coefficient and all others we report in this table are highly statistically significant ($p < 0.001$), with standard errors clustered at the level of siblings. Adding individual controls in column (2) reduces the age 5 life expectancy gap between offspring of cousin and non-cousin parents to 2.2 years.

Column (3) presents our preferred estimates using maternal and paternal fixed effects to control for any factors common to the children of an individual's aunts and uncles. These include a wide range of unobserved economic, cultural and genetic factors common to these close relatives. Including parental fixed effects suggests that first-cousin marriage causes life expectancy at age 5 to decrease by 2.2 years.

To help interpret the magnitude of this result, we can compare it to the relationship between parent and offspring longevity we documented in panel (a) of Figure 1. The reduction in life expectancy for children born to first-cousin parents is as large as the difference observed between children whose parents lived to the median age (71 years) and those who reached only the 10th percentile (54 years).

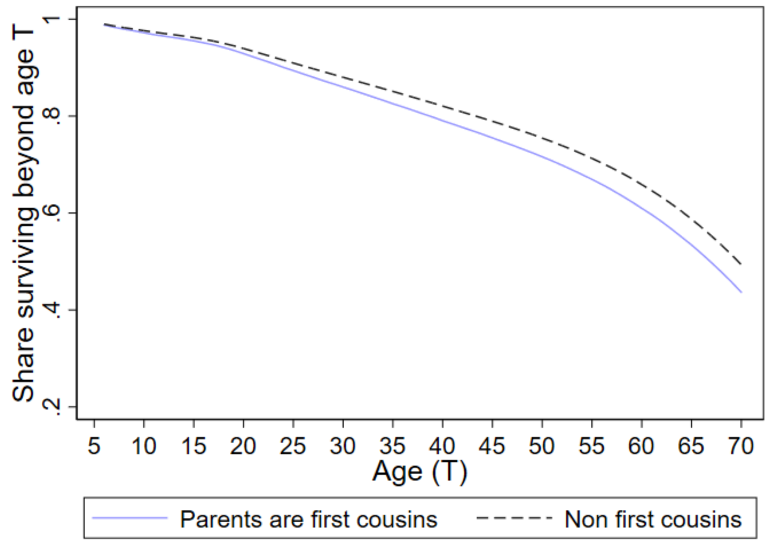
These differences in life expectancy are driven by higher mortality throughout the lifespan. Table 2 also reports results on life expectancy at age 20 (Panel B). Coefficients remain similar, which suggests that the effect of cousin marriage is in large part due to adult mortality. Conditional on living until the age of 20, offspring of first cousins live on average 1.8 fewer years than the offspring of their parents' siblings.

Figure 2 more flexibly documents how survival rates differ for offspring of first cousins across the lifespan. Panel (a) shows raw offspring survival rates conditional on survival to age 5. It suggests that the gap in survival grows gradually over the lifespan, rather than being concentrated in one period. Panel (b) shows regression coefficients and 95% confidence intervals that compare survival rates for offspring of cousin and non-cousin marriages including maternal and paternal fixed effects.[14] It confirms that the pattern observed in Panel (a) holds after controlling for selection. Indeed we see a consistently increasing gap between the survival rates of offspring from cousin and non-cousin marriages. It shows that, conditional on surviving to age 5, having first-cousin parents leads to a four percentage point lower probability of surviving past age fifty. This is a 16% increase in the probability of dying between the ages of five and fifty.

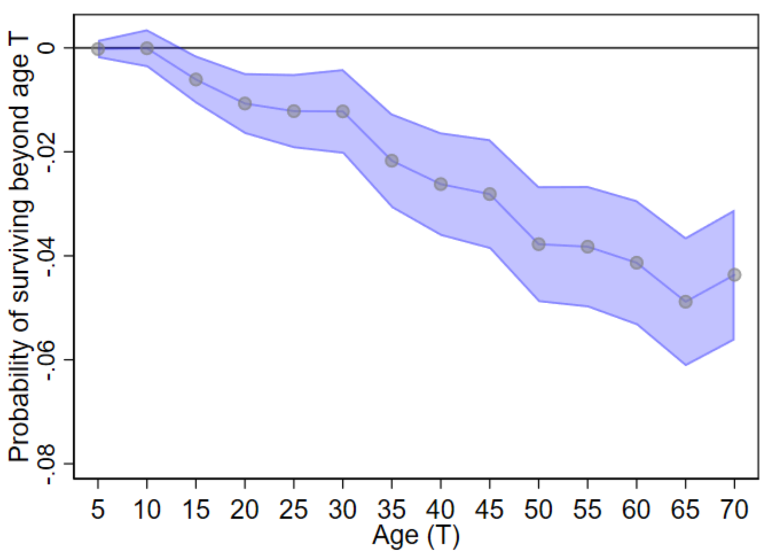
Columns (4) to (7) of Table 2 provide evidence on the sensitivity of our results to data and specification changes. Column (4) replaces the quadratic measures of birth year, maternal age at birth, number of siblings, and birth order with fixed effects for each integer value of these variables. Column (5) adds parental longevity as a control (the mean of father and mother's longevity). Column (6) drops all individuals whose year of death ends in zero. We add this robustness check because of evidence of heaping of death years in

[14]Each estimate in panel (b) is from a regression where the outcome is whether an individual survived to a given age. Otherwise, these regressions are identical to equation (1).

Figure 2: Mean and estimated survival probabilities



(a) Average survival rates



(b) Estimated effect of cousin marriage on survival probabilities

This figure depicts the probabilities of survival for our analysis sample of 5.9 million offspring. Panel (a) shows average survival rates (without controls) for offspring whose parents are first cousins (solid line) and not first cousins (dashed line). Panel (b) shows a series of estimates analogous to the coefficient β from equation (1) estimated using OLS and corresponding 95 % confidence intervals. Estimates include all of the controls and fixed effects used in column (3) of table 2. For each estimate, the outcome variable is equal to 1 if the individual survives beyond age T, and 0 otherwise.

our data, as evidenced by Appendix Figure A.10. Column (7) adds county-by-decade-of-birth fixed effect, since there may be differential sorting into locations with lower life expectancy. Column (8) uses same-surname marriage as an alternative treatment variable, which increases sample size but results in slightly attenuated effects, presumably due to the increase in measurement error.[15] In each column, we find results that are close in magnitude to our baseline specification in column (3).

We conclude this section by highlighting two additional issues with our genealogical data. The first is that not all historical individuals have genealogical profiles and, further, our analysis sample is restricted to offspring for whom genealogical profiles exist for all eight great-grandparents. This means that, as discussed in section 2, our analysis sample is not representative of the entire US population. While the lack of detailed socio-demographic information in our genealogical data does not allow us to fully address such concerns, Appendix Table A.6 offers some evidence for the generalizability of our results. For this table, we reweighted observations to match the entire US-born census population using variables that are available in both the census and the genealogical profiles (sex and region of birth). Estimates in this table are close in magnitude and not statistically distinguishable from our baseline estimates in Table 2.

Finally, historical genealogical data such as ours are not well suited for studying infant or child mortality, since individuals who die young may go unrecorded. This is why our results so far only report life expectancy conditional on survival to age 5. However, the effect of first-cousin marriage on life expectancy at birth is almost certainly greater than its effect on life expectancy at age 5. Appendix Table A.7 shows results comparable to our main results in Table 2, but includes all offspring rather than restricting to those who survived until age 5. The effect of first-cousin marriage increases to 3.2 years of life lost. We note that this result should be interpreted with caution since, as described in section 2, our data are missing a large share of infant deaths. Given the well-documented increase in infant deaths for cousin spouses, the true effect on life expectancy at birth is almost certainly greater than 3.2 years.

[15] See Ghosh et al. (2023) on the use of marital isonymy as a proxy for first-cousin marriage, and a discussion of the associated type 1 and type 2 errors.

5 Conclusion and discussion

This paper uses forty million genealogical profiles to study the effect of first-cousin marriage on the health of offspring. Causal estimates come from comparisons between the offspring of married cousins and the offspring of their siblings. We find that cousin marriage reduces age-5 life expectancy by two years. This difference is the result of increased mortality throughout the lifespan, which highlights the importance of studying adult health outcomes of offspring of cousin marriages. Strikingly, we also find that these effects are stable across 150 years of birth cohorts. Dramatic transformations in the US during this period implies that these effects are not very sensitive to the social or economic environment at a given time.

Are these effects purely genetic? Or are there important social or economic consequences of cousin marriage that lead to offspring having shorter lives? While research in medicine and human biology has focused on genetic effects, recent work in economics suggests there may be important economic consequences to cousin marriage. Notably, Ghosh et al. (2023) finds that cousin marriage leads to lower incomes and reduces rural-to-urban migration, both of which may affect mortality. Cousin marriages may also lead to an increase in the number of children a couple chooses to have, or their timing (e.g., having children at a younger age). Finally, cousin marriage may violate social norms. Each of these may have both biological and economic consequences that affect longevity. Testing various channels falls beyond the scope of this paper.

We believe the results in this paper are informative about health costs in countries where cousin marriage is commonly practiced. During the period of our analysis, the US had mortality rates and income per capita comparable to many of these countries today. Nonetheless, twentieth-century medical advances may have reduced the health costs of first-cousin marriage. Data from the Demographic Health Survey (DHS) suggests this is not the case. In places with high rates of cousin marriage, the survey asks women whether their husband is their first cousin. In every one of these 26 survey-waves, the child mortality rate for women married to a first cousin is higher than for non-cousins, often substantially so. Consistent with our findings using historical US data, this difference seems to be independent of income per capita and of the baseline mortality rate. The increase in child mortality for first cousins in our data is, if anything, smaller than in the DHS samples. This suggests that health costs are not lower today than they were in the nineteenth-century US.

At a first approximation, our findings imply that one seventh of the difference in life expectancy between the US and Pakistan is due to high rates of first-cousin marriage in the latter country.[16] These large effects suggest that cheaper and more widely available genetic counselling and screening may lead to substantial public health improvements in societies where cousin marriage is common.

[16]Using 2020 World Bank estimates of life expectancy of 77.3 and 66.3 years for the US and Pakistan, respectively. We use first-cousin marriage rates of 0% and 50%, and a reduction of 3.2 years for offspring of cousin marriages.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson (2012) "Europe's tired, poor, huddled masses: Self-selection and economic outcomes in the age of mass migration," *American Economic Review*, 102 (5), 1832–1856.

Almond, Douglas and Bhashkar Mazumder (2011) "Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy," *American Economic Journal: Applied Economics*, 3 (4), 56–85.

Atkin, David (2016) "The caloric costs of culture: Evidence from Indian migrants," *American Economic Review*, 106 (4), 1144–1181.

Bahrami-Rad, Duman (2021) "Keeping it in the family: Female inheritance, inmarriage, and the status of women," *Journal of Development Economics*, 153, 102714.

Bener, Abdulbari, Hanadi R El Ayoubi, Lotfi Chouchane, Awab I Ali, Aisha Al-Kubaisi, Haya Al-Sulaiti, and Ahmad S Teebi (2009) "Impact of consanguinity on cancer in a highly endogamous population," *Asian Pac J Cancer Prev*, 10 (1), 35–40.

Bener, Abdulbari, Hanadi Rafii El Ayoubi, Awab Ibrahim Ali, Aisha Al-Kubaisi, and Haya Al-Sulaiti (2010) "Does consanguinity lead to decreased incidence of breast cancer?" *Cancer Epidemiology*, 34 (4), 413–418.

Bener, Abdulbari, Rafat Hussain, and Ahmad S Teebi (2007) "Consanguineous marriages and their effects on common adult diseases: studies from an endogamous population," *Medical Principles and Practice*, 16 (4), 262–267.

Bennett, Robin L, Neeraja R Malleda, Peter H Byers, Robert D Steiner, and Kimberly M Barr (2021) "Genetic counseling and screening of consanguineous couples and their offspring practice resource: Focused Revision," *Journal of Genetic Counseling*, 30 (5), 1354–1357.

Bennett, Robin L, Arno G Motulsky, Alan Bittles et al. (2002) "Genetic counseling and screening of consanguineous couples and their offspring: Recommendations of the National Society of Genetic Counselors," *Journal of Genetic Counseling*, 11 (2), 97–119.

Bittles, Alan H and Michael L Black (2010) "Consanguinity, human evolution, and complex diseases," *Proceedings of the National Academy of Sciences*, 107, 1779–1786.

Bittles, Alan H and James V Neel (1994) "The costs of human inbreeding and their implications for variations at the DNA level," *Nature genetics*, 8 (2), 117–121.

Black, Sandra E, Neil Duzett, Adriana Lleras-Muney, Nolan G Pope, and Joseph Price (2023) "Intergenerational Correlations in Longevity," Working Paper 31034, National Bureau of Economic Research, 10.3386/w31034.

Blanc, Guillaume (2023) "Crowdsourced Genealogies," working paper, University of Manchester.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley E.B. Wilbert (2023) "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project," Working Paper 31671, National Bureau of Economic Research, 10.3386/w31671.

Collins, William J and Marianne H Wanamaker (2014) "Selection and economic gains in the great migration of African Americans: new evidence from linked census data," *American Economic Journal: Applied Economics*, 6 (1), 220–252.

Corno, Lucia, Eliana La Ferrara, and Alessandra Voena (2020) "Female Genital Cutting and the Slave Trade," CEPR Discussion Paper DP15577, CEPR, `https://ssrn.com/abstract=3753982`.

Correia, Sergio (2015) "Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix," working paper, Duke University, `http://scorreia.com/research/singletons.pdf`.

Currie, J et al. (2009) "Healthy, wealthy, and wise: socioeconomic status, poor health in childhood, and human capital development.," *Journal of Economic Literature*, 47 (1), 87–117.

Denic, Srdjan, Chris Frampton, and M Gary Nicholls (2007) "Risk of cancer in an inbred population," *Cancer Detection and Prevention*, 31 (4), 263–269.

Fernández, Raquel (2011) "Does culture matter?" *Handbook of Social Economics*, 1, 481–510.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams (2021) "Place-based drivers of mortality: Evidence from migration," *American Economic Review*, 111 (8), 2697–2735.

Ghosh, Arkadev, Sam Il Myoung Hwang, and Munir Squires (2023) "Economic Consequences of Kinship: Evidence From U.S. Bans on Cousin Marriage," *The Quarterly Journal of Economics*, 138 (4), 2559–2606.

Gibson, Campbell and Kay Jung (2002) "Historical Census Statistics on Population Totals by Race, 1790 to 1990, and by Hispanic origin, 1970 to 1990, for the United States, Regions, Divisions, and States," Working Paper POP-WP056, US Census Bureau.

Gilani, Ghausia Masood, Shahid Kamal, and Syed Aamir Masood Gilani (2006) "Risk factors for breast cancer for women in Punjab, Pakistan: Results from a case-control study," *Pakistan Journal of Statistics and Operation Research*, 17–26.

Giuliano, Paola and Nathan Nunn (2021) "Understanding cultural persistence and change," *The Review of Economic Studies*, 88 (4), 1541–1581.

Hacker, J David (2010) "Decennial life tables for the white population of the United States, 1790–1900," *Historical methods*, 43 (2), 45–79.

Haines, Michael R. and Inter-university Consortium for Political and Social Research (2010) "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002," 10.3886/ICPSR02896.v3.

Hamamy, Hanan, Stylianos E Antonarakis, Luigi Luca Cavalli-Sforza et al. (2011) "Consanguineous marriages, pearls and perils: Geneva international consanguinity workshop report," *Genetics in Medicine*, 13 (9), 841–847.

Helgason, Agnar, Snæbjörn Pálsson, Daníel F Gudbjartsson, Thórdur Kristjánsson, and Kári Stefánsson (2008) "An association between the kinship and fertility of human couples," *Science*, 319 (5864), 813–816.

Hwang, Sam Il Myoung and Munir Squires (2024a) "Linked samples and measurement error in historical US census data," *Explorations in Economic History*, 101579.

——— (2024b) "Linked samples and measurement error in historical US census data," *Explorations in Economic History*, 101579.

Kaplanis, Joanna, Assaf Gordon, Tal Shor et al. (2018) "Quantitative analysis of population-scale family trees with millions of relatives," *Science*, 360 (6385), 171–175.

Kreisman, Daniel and Jonathan Smith (2023) "Distinctively Black names and educational outcomes," *Journal of Political Economy*, 131 (4), 877–897.

Lévi-Strauss, Claude (1969) *The Elementary Structures of Kinship*, Boston: Beacon Press, Original French edition published in 1949.

Liede, Alexander, Imtiaz A. Malik, Zeba Aziz, Patricia de los Rios, Elaine Kwan, and Steven A. Narod (2002) "Contribution of BRCA1 and BRCA2 Mutations to Breast and Ovarian Cancer in Pakistan," *American Journal of Human Genetics*, 71 (3), 595–606.

Long, Jason and Joseph Ferrie (2013) "Intergenerational occupational mobility in Great Britain and the United States since 1850," *American Economic Review*, 103 (4), 1109–1137.

Lowes, Sara and Eduardo Montero (2021) "The legacy of colonial medicine in Central Africa," *American Economic Review*, 111 (4), 1284–1314.

Lu, Frances and Tom Vogl (2023) "Intergenerational persistence in child mortality," *American Economic Review: Insights*, 5 (1), 93–109.

McWhirter, Rebekah E, Ruth McQuillan, Elizabeth Visser, Carl Counsell, and James F Wilson (2012) "Genome-wide homozygosity and multiple sclerosis in Orkney and Shetland Islanders," *European Journal of Human Genetics*, 20 (2), 198–202.

Mobarak, A Mushfiq, Theresa Chaudhry, Julia Brown et al. (2019) "Estimating the health and socioeconomic effects of cousin marriage in South Asia," *Journal of Biosocial Science*, 51 (3), 418–435.

Mobarak, Ahmed Mushfiq, Randall Kuhn, and Christina Peters (2013) "Consanguinity and other marriage market effects of a wealth shock in Bangladesh," *Demography*, 50 (5), 1845–1871.

Paul, Diane B and Hamish G Spencer (2008) ""It's ok, we're not cousins by blood": the cousin marriage controversy in historical perspective," *PLoS Biology*, 6 (12), e320.

Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley (2021) "Combining family history and machine learning to link historical records: The Census Tree data set," *Explorations in Economic History*, 80, 101391.

Rudan, Igor, D Rudan, H Campbell et al. (2003b) "Inbreeding and risk of late onset complex disease," *Journal of Medical Genetics*, 40 (12), 925–932.

Rudan, Igor, Nina Smolej-Narancic, Harry Campbell, Andrew Carothers, Alan Wright, Branka Janicijevic, and Pavao Rudan (2003a) "Inbreeding and the genetic complexity of human hypertension," *Genetics*, 163 (3), 1011–1021.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler (2023) "Version 13.0 1 percent random sample of U.S. Federal Census, 1850-1930," `https://doi.org/10.18128/D010.V13.0`.

Strauss, John and Duncan Thomas (2007) "Health over the life course," *Handbook of Development Economics*, 4, 3375–3474.

The Economist (2024) "All in the family," *The Economist*, 450 (9384), 31–32, `https://www.economist.com/united-states/2024/02/15/cousin-marriage-is-probably-fine-in-most-cases`, Accessed: 2024-02-22.

US Census Bureau (2021) "Historical Population Change Data (1910-2020)," April, `https://www.census.gov/data/tables/time-series/dec/popchange-data-text.html`, Accessed: 2023-10-13.

Ward, Zachary (2022) "Internal Migration, Education, and Intergenerational Mobility Evidence from American History," *Journal of Human Resources*, 57 (6), 1981–2011.

A Online Appendix: Additional Tables and Figures

Table A.1: Construction of analysis sample

	(1) Total Dropped	(2) Percent Dropped	(3) Remaining Observations
Sex	60,501	0.15	40,514,188
Nonmissing birth year	581,967	1.44	39,932,221
Nonmissing death year	10,903,538	27.31	29,028,683
Nonmissing great-grandparents	21,618,119	74.47	7,410,564
Nonmissing maternal age at birth	29,440	0.40	7,381,124
Longevity between 0 and 98	49,568	0.67	7,331,556
Birth year between 1750 and 1920	665,908	9.08	6,665,648
Survived to age 5	608,335	9.13	6,057,313
Singletons	142,789	2.36	5,914,524

This table shows how we create our final analysis sample of 5.9 million offspring from over 40 million genealogical records. Each row shows the number of observations remaining after we drop those for which a specific variable is missing. Singletons are groups with only one observation. These are relevant for specifications in which we include mother and father siblings fixed effects. See Correia (2015) for a more detailed description of singletons.

Table A.2: Descriptive statistics - analysis sample

Analysis sample: Individuals with non-missing great-grandparents

	(1) Parents are first cousins	(2) Non-cousin	(3) Difference
Longevity conditional on surviving to age 5	60.97	63.73	-2.77
	[22.58]	[22.30]	(0.06)
Parent Longevity	69.13	69.66	-0.53
	[11.90]	[11.59]	(0.03)
Year of Birth	1,842.84	1,848.36	-5.51
	[32.35]	[34.25]	(0.09)
Mother's Age at Birth	30.14	29.79	0.35
	[7.02]	[6.97]	(0.02)
Female	0.47	0.47	-0.00
			(0.00)
Number of brothers	4.21	4.16	0.05
	[2.21]	[2.22]	(0.01)
Number of sisters	3.91	3.85	0.07
	[2.14]	[2.15]	(0.01)
Birth order	4.39	4.33	0.06
	[2.75]	[2.73]	(0.01)
Observations	148,686	5,765,838	5,914,524
Percent	2.51	97.49	100

Each observation is an offspring in the analysis sample. This table shows the mean of each variable we use in our preferred specification in table 2. We require children to survive until at least age 5 to be included in the analysis sample we describe here. Column (1) shows means for children whose parents are first cousins. Column (2) shows means for children whose parents are not first cousins. Column (3) shows the difference between columns (1) and (2). Parental longevity is substantially higher than child longevity since it only includes individuals who have children (and hence survived to reproductive age). Variable descriptions are in section 3. Standard deviations are in square brackets. Standard errors are in parentheses.

Table A.3: Comparison of socio-demographic characteristics between key samples (1850)

	(1) 1% Census sample (IPUMS)	(2) Subset of (1) with genealogical profile	(3) Intersection of (2) and (4)	(4) Genealogical profiles in analysis sample
Panel A: Variables available in census records and genealogical profiles				
Female	.49	.49	.46	.46
Age in 1850	22.34	21.41	23.33	24.15
	[17.56]	[17.51]	[18.65]	[19.27]
Born in Northeast	.41	.43	.52	.51
Born in Midwest	.15	.18	.19	.18
Born in South	.32	.34	.28	.29
Born in West	0	0	0	0
Foreign-born	.11	.04	0	.02
Panel B: Variables available only in census records				
Non-white	.02	0	0	
Related to head	.91	.98	.98	
Literate	.89	.9	.95	
White-collar	.1	.09	.11	
Farmer	.45	.59	.62	
Skilled	.26	.2	.18	
Unskilled	.19	.12	.09	
Live in urban area	.17	.1	.08	
Live on a farm	.53	.63	.69	
Value of real estate	249.92	296.26	470.27	
	[2978.47]	[3065.54]	[2867.36]	
Panel C: Variables available only in genealogical profiles				
Longevity conditional on surviving to age 5		65.11	65.65	66.09
		[19.99]	[19.83]	[19.65]
Mother's Age at Birth		29.02	29.22	29.33
		[7.03]	[6.94]	[6.95]
Number of brothers		4.08	4.37	4.34
		[2.25]	[2.22]	[2.25]
Number of sisters		3.76	3.95	3.98
		[2.13]	[2.13]	[2.16]
Birth order		3.94	4.15	4.2
		[2.71]	[2.69]	[2.72]
Sibling sex ratio		.48	.47	.48
		[.2]	[.19]	[.2]
Observations	197,796	109,825	15,592	2,257,851

Note: This table compares the characteristics of individuals from four samples who were alive in 1850 (born pre-1850 and died post-1850). Column (1) corresponds to the 1850 U.S. Federal Census IPUMS 1% sample (Ruggles et al., 2023); column (2) to the subsample of (1) linked to a genealogical profile on FamilySearch, as created in Hwang and Squires (2024a); column (3) to the subsample of (2) that overlaps with our analysis sample; and (4) to our analysis sample. The variables in Panel A are available both in the census sample and our analysis sample, while Panels B and C contain variables available in one dataset but not the other. The numbers in the brackets represent standard deviations. We group occupations into white-collar, farmer, skilled, and unskilled following the categories used in Long and Ferrie (2013).

Table A.4: Descriptive statistics - parent sample

Parent sample: Parents of analysis sample

	(1) Married to first cousin	(2) Not married to first cousin	(3) Difference
Longevity	67.38	67.95	-0.57
	[16.96]	[16.89]	(0.09)
Year of Birth	1,812.84	1,817.55	-4.71
	[25.96]	[28.06]	(0.15)
Mother's Age at Birth	29.39	30.07	-0.68
	[6.87]	[7.24]	(0.04)
Number of Children	6.80	6.54	0.27
	[3.42]	[3.41]	(0.02)
Female	0.48	0.49	-0.01
			(0.00)
Number of brothers	2.62	2.44	0.19
	[1.74]	[1.67]	(0.01)
Number of sisters	2.35	2.28	0.07
	[1.57]	[1.53]	(0.01)
Birth order	2.87	2.87	-0.00
	[1.80]	[1.80]	(0.01)
Observations	35,366	1,219,265	1,254,631
Percent	2.82	97.18	100

Each observation is a parent of one of the offspring in our analysis sample. This table shows the mean of each variable we use in our preferred specification in table 1. Column (1) shows means for parents who are married to their first cousins. Column (2) shows means for parents who are not married to their first cousins. Column (3) shows the difference between columns (1) and (2). Variable descriptions are in section 3. Standard deviations are in square brackets. Standard errors are in parentheses.

Table A.5: Comparison of Socio-Demographic Characteristics Between First-Cousin Couples and Non-First-Cousin Couples

	Married to first cousin		Not married to first cousin		Difference	
	Obs.	Mean	Obs.	Mean	Raw	Sib. FE
Female	142	.51	6,211	.48	.03 (.04)	.02 (.04)
Born in Northeast	142	.24	6,211	.21	.03 (.03)	.02 (.01)
Born in Midwest	142	.11	6,211	.32	-.21 (.04)	-.02 (.02)
Born in South	142	.65	6,211	.45	.2 (.04)	0 (.01)
Born in West	142	.01	6,211	.02	-.01 (.01)	0 (0)
Foreign-born	142	0	6,211	.01	-.01 (.01)	0 (.01)
Non-white	142	0	6,211	0	0 (0)	0 (0)
Related to head	142	.99	6,211	.99	0 (.01)	-.01 (.01)
Literate	77	.84	3,110	.92	-.08 (.03)	-.04 (.04)
White-collar	45	.04	1,832	.08	-.04 (.04)	.12 (.08)
Farmer	45	.71	1,832	.59	.12 (.07)	-.18 (.12)
Skilled	45	.18	1,832	.11	.07 (.05)	0 (.09)
Unskilled	45	.07	1,832	.22	-.15 (.06)	.06 (.1)
Live in urban area	142	.05	6,211	.05	0 (.02)	-.01 (.02)
Live on a farm	142	.81	6,211	.79	.02 (.03)	.02 (.03)
Value of real estate	111	232.43 [688.89]	4,776	336.32 [2411.93]	-103.89 (229.17)	-39.98 (296.81)
Value of asset	51	123.8 [371.5]	2,145	290.08 [1450.2]	-166.28 (203.27)	11.78 (253.77)

This table presents the results of a t-test of differences in socio-demographic characteristics between couples married to first cousins and those who are not. We restrict our sample to those who satisfy the following two conditions: (1) those whom we can link to the 1850-1930 IPUMS 1% samples (Ruggles et al., 2023), the linkage of which is created in Hwang and Squires (2024b); and (2) those who have a sibling that is linked as well. When a person is linked to the IPUMS samples multiple times (2.3 to 4.6 percent, depending on the characteristics), we use the average of the socio-demographic characteristics. The number of observations differs across rows due to differences in the universe for the census questions or differences in the share of missing values. The column labeled "Raw" presents the raw difference in the sample means and the standard errors. The column labeled "Sib. FE" displays the differences in the sample mean between two groups after including sibling fixed effects.

Table A.6: The effect of cousin marriage on longevity with observations reweighted to match sex and birth region in the U.S. Census

	(1) Raw	(2) Controls	(3) Parental fixed effects
Panel A: Life expectancy at age 5			
Parents are first cousins	-2.86	-2.18	-1.94
	(0.10)	(0.10)	(0.56)
Control mean	64.65	64.65	64.65
Observations (thousands)	3,188	3,188	3,188
Panel B: Life expectancy at age 20			
Parents are first cousins	-2.35	-1.83	-1.61
	(0.08)	(0.08)	(0.00)
Control mean	67.98	67.98	67.98
Observations (thousands)	2,980	2,980	2,980
Individual controls	No	Yes	Yes
Paternal fixed effects	No	No	Yes
Maternal fixed effects	No	No	Yes

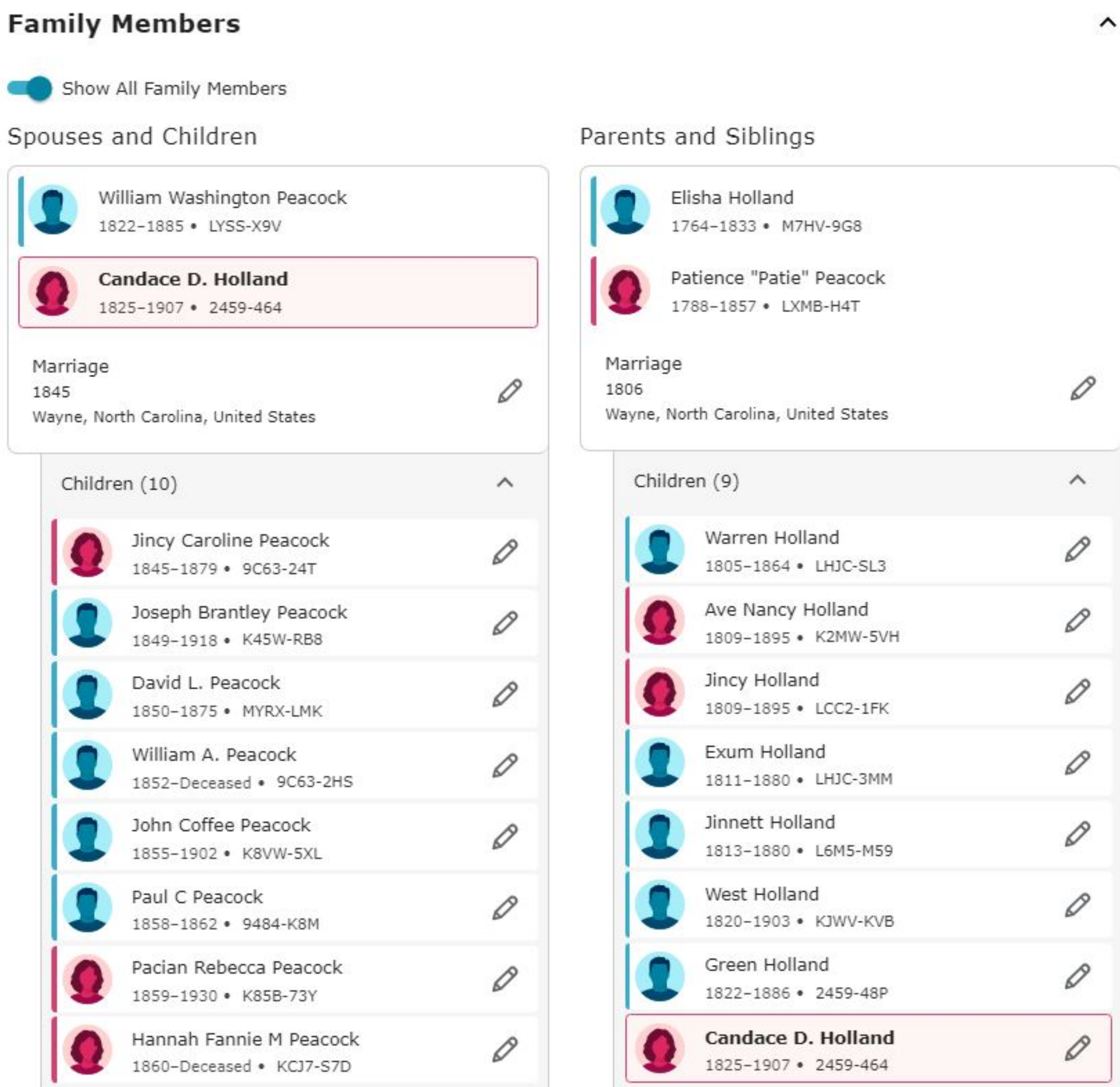
Note: This table shows estimates for the coefficient β from equation (1). Each observation is an offspring in the analysis sample. The outcome is the child's longevity (year of death minus year of birth), conditional on surviving to a specified age. Each regression weights members of each decadal birth cohort from 1840 to 1910 to match the sex and birth region of whites (Northeast, Midwest, South, and West) in the census closest to their birth. For example, the 1840 birth cohort (those born between 1840 and 1849) is weighted so that the weighted share of each sex $\times$ birth region of whites matches the corresponding share in the full-count 1850 census. We exclude the 1880 birth cohort because the 1890 full-count census is not available. Column (1) coefficients are simply the weighted difference in mean longevity between the children of first cousins and the children of non-first cousins. Column (2) adds controls for birth year, sex, maternal age at birth, number of sisters, number of brothers, the sex ratio of siblings, and birth order, as described in section 3. Column (3) adds mother's siblings fixed effects and father's siblings fixed effects. Standard errors are clustered at the level of the individual and their siblings. We restrict our sample to those born in the U.S.

Table A.7: Life expectancy at birth

	Baseline specifications			Robustness checks				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Raw	Controls	Parental fixed effects	Flexible controls	Parent longevity	Age heaping	County-decade FE	Same surname
Life expectancy at birth								
Parents are first cousins	-3.10	-2.60	-3.18	-3.18	-3.08	-2.82	-3.05	-2.27
	(0.09)	(0.09)	(0.33)	(0.33)	(0.35)	(0.37)	(0.42)	(0.22)
Control mean	58.02	58.02	58.02	58.02	58.03	58.22	58.40	57.97
Observations (thousands)	6,539	6,539	6,539	6,539	6,461	5,744	5,069	13,978
Individual controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Paternal fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Maternal fixed effects	No	No	Yes	Yes	Yes	Yes	Yes	Yes

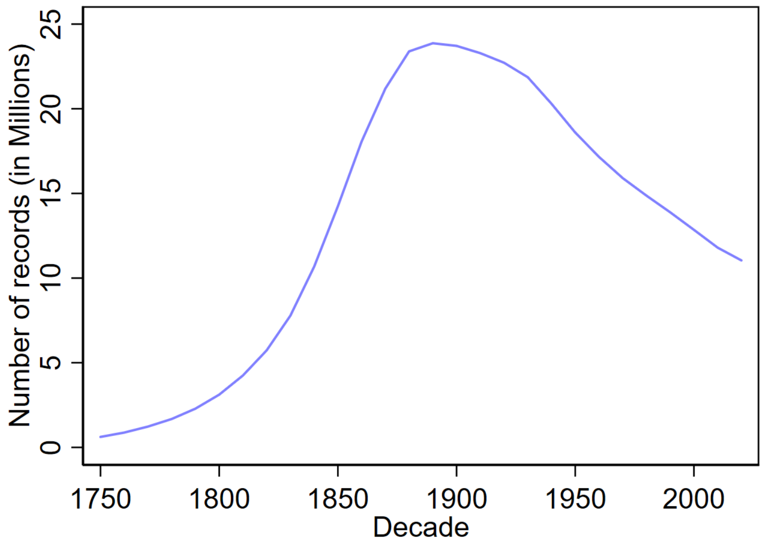
This table shows estimates for the coefficient β from equation (1) estimated using OLS. Each observation is a child in the analysis sample. The outcome is the child's longevity (year of death minus year of birth). In this table we do not require the child to have survived to a certain age. Column (1) coefficients are simply the difference in mean longevity between the children of first cousins and the children of non-first cousins. Column (2) adds controls for birth year, sex, maternal age at birth, number of sisters, number of brothers, the sex ratio of siblings, and birth order, as described in section 3. Column (3) adds mother's siblings fixed effects and father's siblings fixed effects. Column (4) replaces the quadratic controls with sets of fixed effects for each integer value. Column (5) controls for parent longevity. Column (6) drops all individuals with death dates ending in 0. Column (7) adds county-by-decade-of-birth fixed effects and removes controls for birth year. In column (8), the treatment variable is equal to 1 if the child's parents have an identical surname, and 0 otherwise. We use the mother's father's surname instead of her own surname to account for the fact that she may have taken her husband's name in marriage. Standard errors are clustered at the level of the individual and their siblings.

Figure A.1: Example FamilySearch Profile

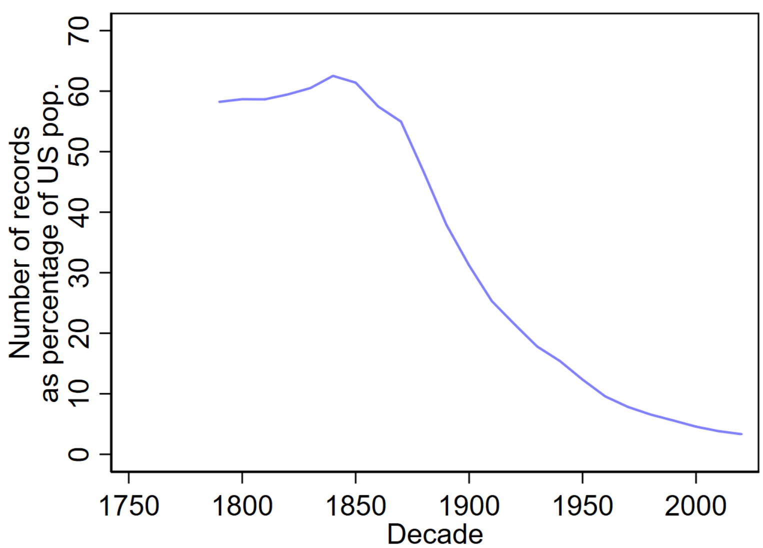


Note: This figure depicts a typical FamilySearch profile (that of Candace D. Holland). Place of birth is not shown in this example.

Figure A.2: Record coverage



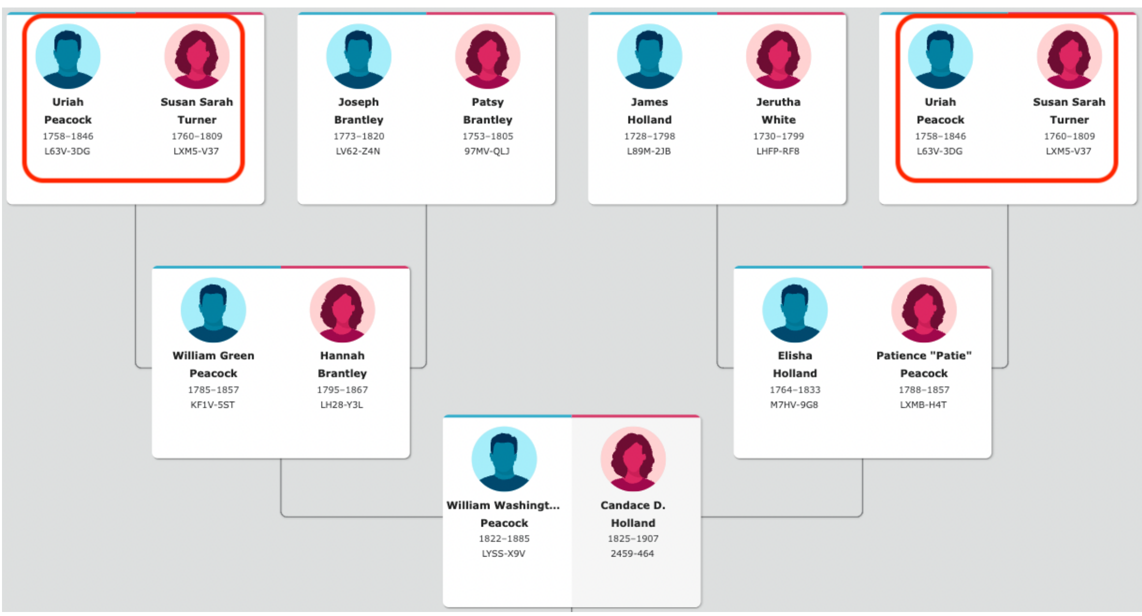
(a) Number of individuals in our dataset alive per decade



(b) Number of individuals alive as percentage of US population

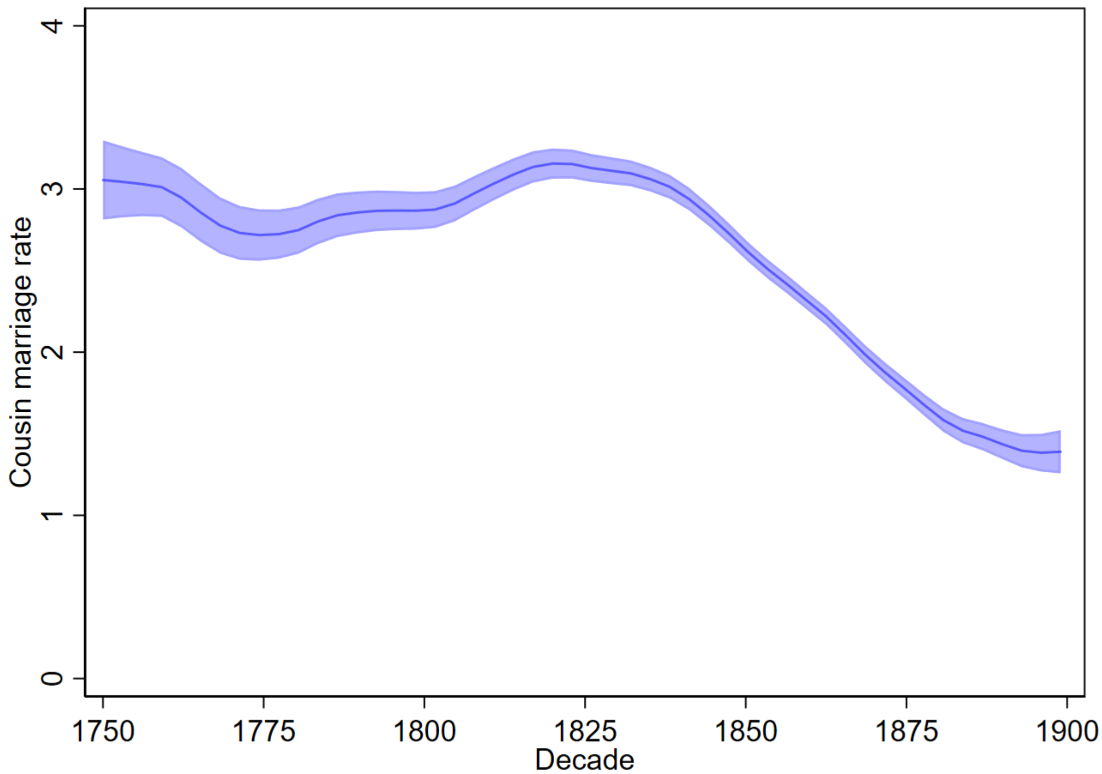
Note: Panel (a) shows the total number of records in our full dataset of 40 million individuals. An individual is counted if they were alive at any point in a given decade. Panel (b) shows these records as a percentage of the US population at the time. US population estimates come from US Census Bureau (2021) for years 2000-2020 and Gibson and Jung (2002) for all other years.

Figure A.3: Genealogical profile of first cousin spouses



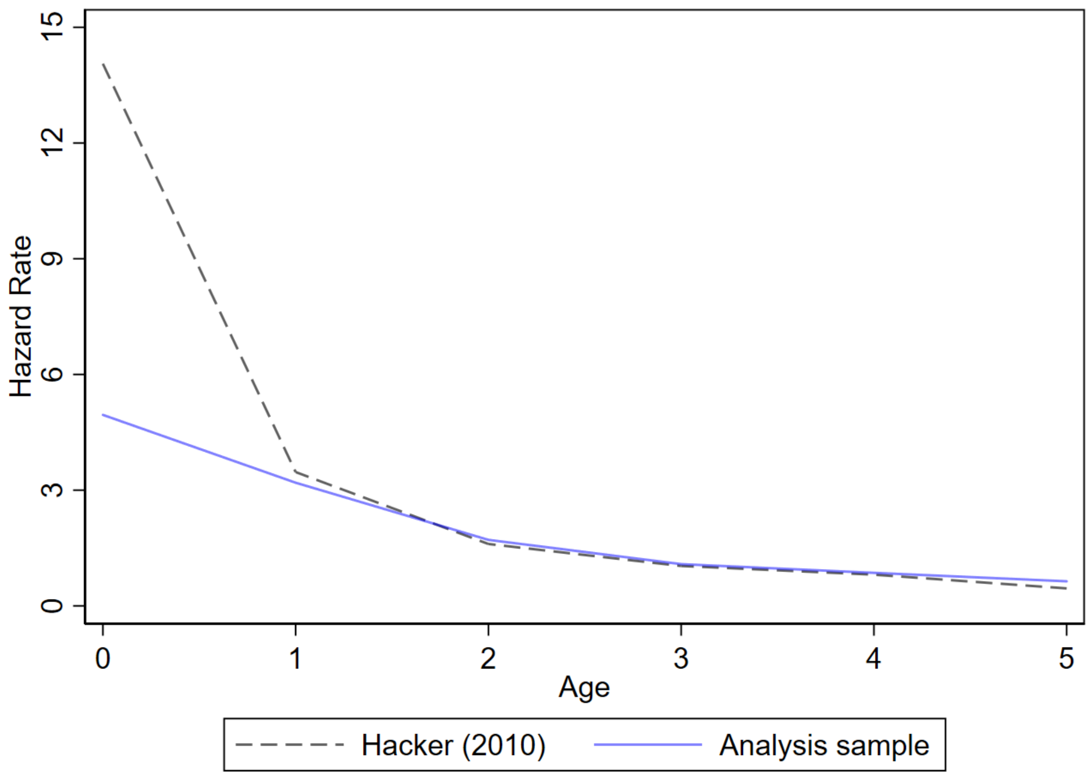
Note: This figure taken from FamilySearch shows the parents and grandparents of spouses (William and Candace) whose names and vital dates are in the bottom row. The husband's father and the wife's mother are siblings. This can be seen by observing the overlapping set of grandparents in the top row of profiles, highlighted in red.

Figure A.4: Cousin marriage rates over time



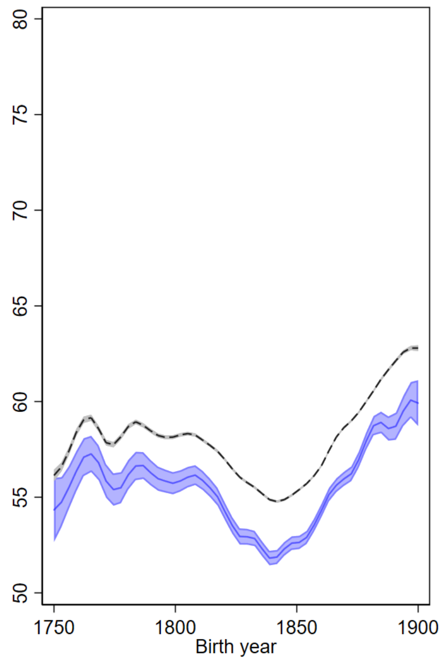
This figure depicts the share of marriages in our analysis sample of 5.9 million offspring that are between first cousins. As a proxy for year of marriage, this figure uses the year of birth of the first child born of a given union. The rate is computed by taking the number of first-born children with first-cousin parents in a given decade divided by the total number of first-born children born in that decade.

Figure A.5: Mortality rates, ages 0-5

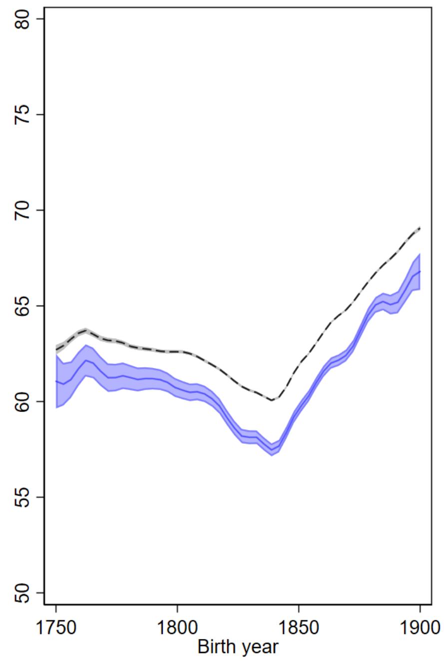


This figure depicts hazard rates for the male offspring in our analysis sample, including those who died before age 5, who died between 1880-1889. We define hazard as the percentage of individuals who die at a given age, conditional on surviving to that age. Historical longevity estimates depicted by the dashed line are from Table 8 of Hacker (2010). The paper argues that female data from this period are estimated with more error, so we use his measure for male mortality from 1880-89.

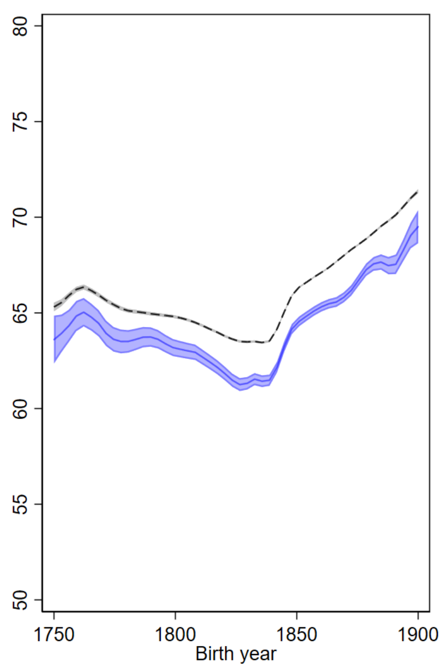
Figure A.6: Life expectancy by birth cohort (at birth and at age 5, 20, and 60)



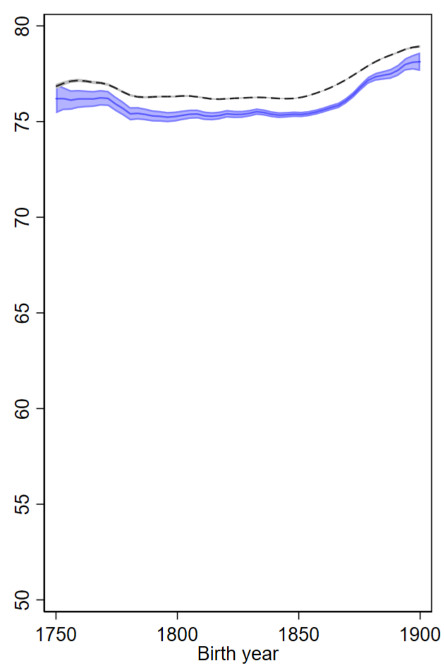
(a) At birth



(b) Life expectancy at age 5



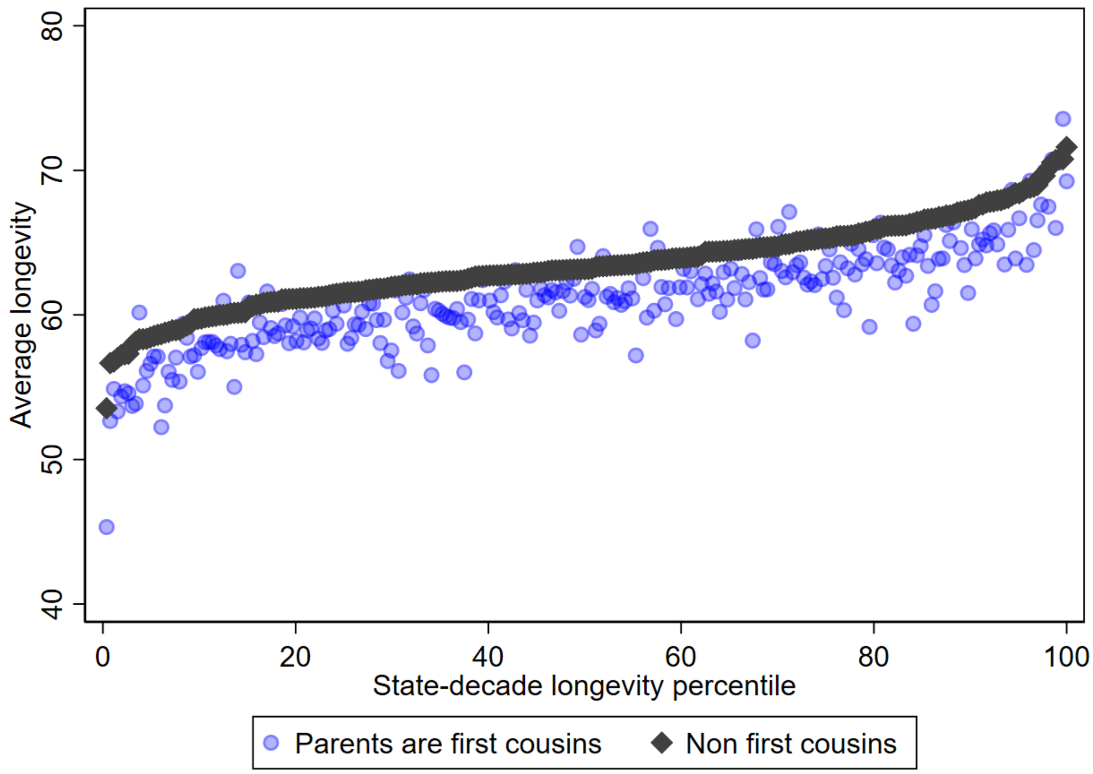
(c) Life expectancy at age 20



(d) Conditional on surviving to age 60

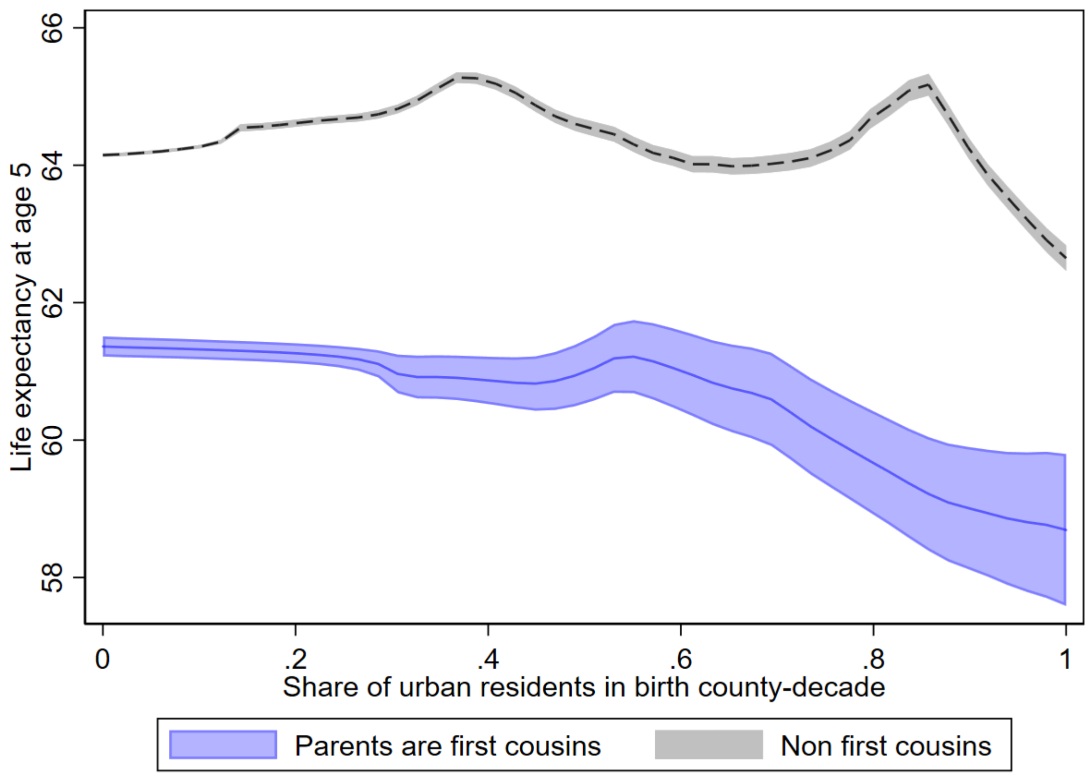
This figure depicts life expectancy of offspring in our analysis sample conditional on surviving to a specified age. Offspring of first cousins are represented by solid blue lines and offspring of non first cousins are represented by dashed gray lines. Panel (a) is a local polynomial regression of life expectancy at birth on birth year. Panels (b), (c), and (d) are local polynomial regressions of life expectancy at age 5, 20, and 60, respectively, on birth year. The shaded areas represent 95% confidence intervals.

Figure A.7: Life expectancy at age 5 by state-decade of birth



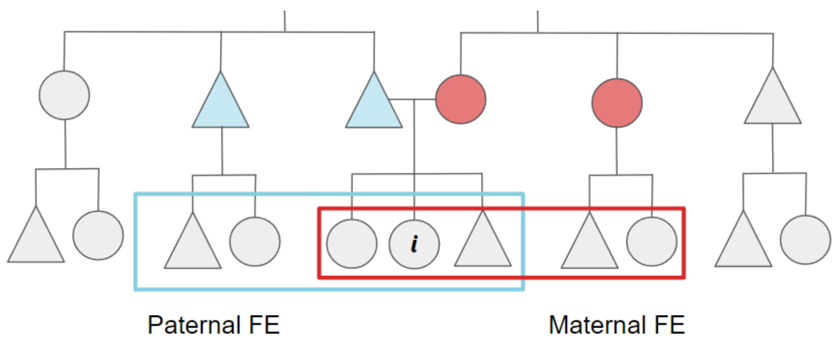
This figure depicts the average longevity at age 5 by state of birth and decade (without controls) for the 5.7 million offspring in our analysis sample for which state of birth is available. Each point is a state-decade pair. Data are sorted by the mean longevity of individuals whose parents are not first cousins in a state-decade pair.

Figure A.8: Life expectancy at age 5 and share of urban residents in birth county-decade



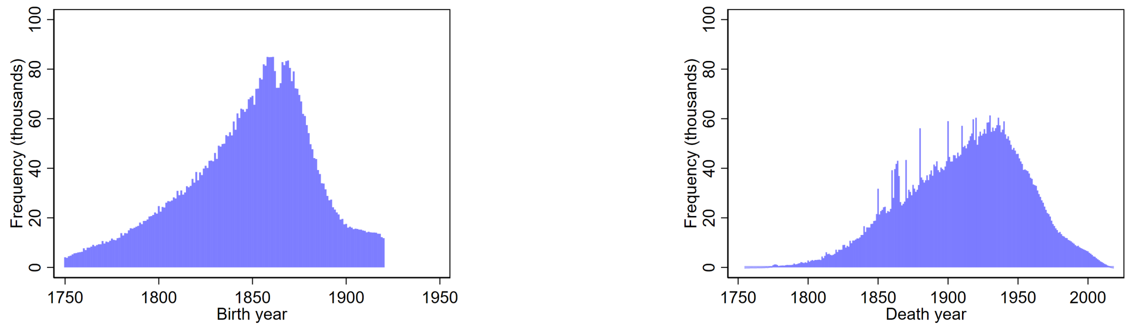
Note: This figure describes the correlation between life expectancy and the share of urban residents in one's birth county-decade. The sample for this figure consists of 4.2 million offspring in our analysis sample whose birth county is observed. The two curves shown in the figure are local polynomial regressions of life expectancy at age 5 on the share of urban residents in one's birth county-decade. The shaded areas represent 95% confidence intervals. The data on county-decade-level shares of urban residents come from Haines and Inter-university Consortium for Political and Social Research (2010).

Figure A.9: Empirical design

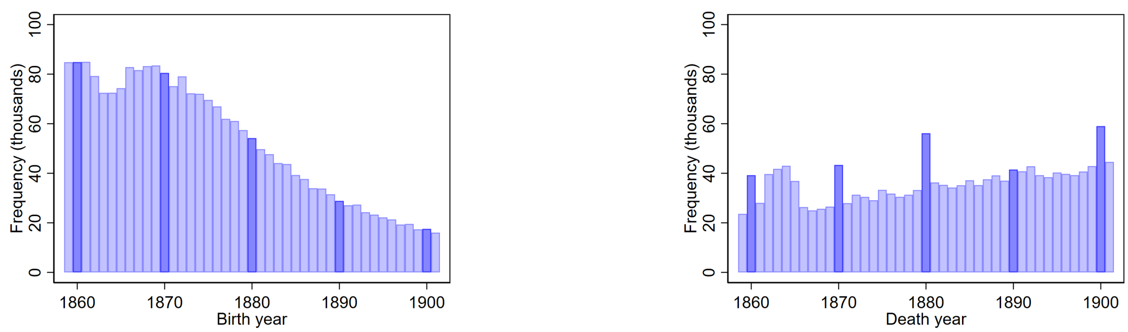


Notes: This figure visualizes the *Maternal* and *Paternal* fixed effects through two generations of related males (triangles) and females (circles). The bottom row represents the 'offspring' of married cousins or non-cousins, and represent the observations in our analysis. The blue and red rectangles represent the maternal and paternal fixed effects that apply to an individual i. These include the maternal and paternal (parallel) cousins of that focal individual i, corresponding to the children of their mother's sisters (red) and their father's brothers (blue).

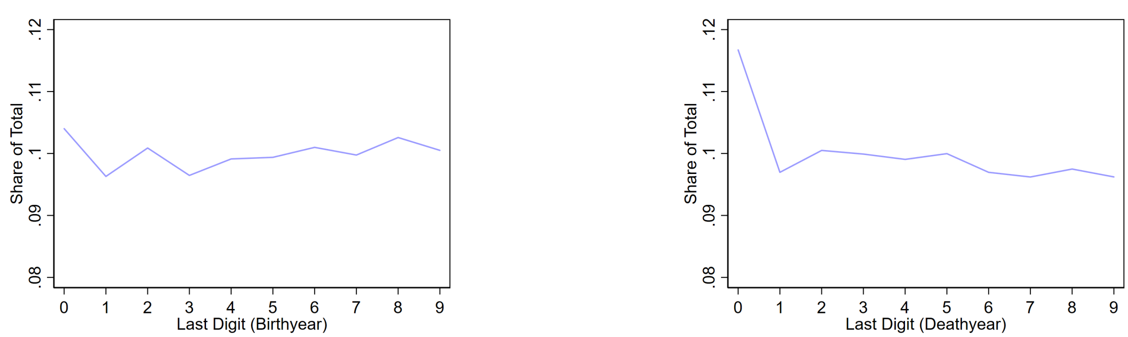
Figure A.10: Data quality: birth and death year heaping



(a) Distribution of vital years in analysis sample



(b) Heaping in decadal census years



(c) Frequency of last digit in vital records

This figure describes age heaping in our analysis sample of 5.9 million offspring. Panels (a) and (b) depict the frequency (in thousands) of birth years and death years. Panels (c) and (d) depict particular segments of (a) and (b), respectively. Census years (ending in zero) are highlighted in a darker shade. Note that individual records for the 1890 census were lost in a fire and hence are not available. Panels (e) and (f) depict the percentage of birth years and death years ending in each digit.