I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16953

Birth Order in the Very Long-Run: Estimating Firstborn Premiums between 1850 and 1940

Angela Cools
Jared Grooms
Krzysztof Karbownik
Siobhan O'Keefe
Joseph Price
Anthony Wray

APRIL 2024

DISCUSSION PAPER SERIES

IZA DP No. 16953

Birth Order in the Very Long-Run: Estimating Firstborn Premiums between 1850 and 1940

Angela Cools
Davidson College

Jared Grooms
Brigham Young University

Krzysztof Karbownik
Emory University, NBER and IZA

Siobhan O'Keefe
Davidson College

Joseph Price
Brigham Young University, NBER and IZA

Anthony Wray
University of Southern Denmark

APRIL 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Birth Order in the Very Long-Run: Estimating Firstborn Premiums between 1850 and 1940*

The nineteenth-century American family experienced tremendous demographic, economic, and institutional changes. By using birth order effects as a proxy for family environment, and linked census data on men born between 1835 and 1910, we study how the family's role in human capital production evolved over this period. We find firstborn premiums for occupational outcomes, marriage, and fertility that are similar across census waves. Our results indicate that the returns to investments in the family environment were stable over a long period.

JEL Classification: J13, J62, N30

Keywords: birth order, parental investments, occupational outcomes, intergenerational mobility, marriage, fertility

Corresponding author:
Krzysztof Karbownik
Department of Economics
Emory University
Rich Memorial Building
1602 Fishburne Drive
30322, Atlanta, Georgia
USA

E-mail: krzysztof.karbownik@emory.edu

* We thank Carver Coleman and J.P. Lefgren for excellent research assistance, Scott Barkowski, Steven Rivkin, and Marian Vidal-Fernandez for helpful comments, and seminar and conference participants at Emory University, Kennesaw State University, the University of Melbourne, the University of New South Wales, and the University of Sydney, the Family and Education Workshop in Uppsala, Sweden, the Social Science History Association Annual Meeting in Chicago, IL, the Society for Economics of the Household Annual Meeting in London, England, the Southeastern Micro Labor Conference in Columbia, SC, and the Southern Economic Association Annual Meeting in Houston, TX.

1 Introduction

Economists and policymakers have long focused on understanding the production function for human capital and now give more emphasis to the role of family environment as a key input (Cunha et al. 2010; Almond and Currie 2011; Almond et al. 2018). Since children spend most of their early years interacting with parents and siblings rather than peers or teachers, it seems intuitive that household conditions contribute meaningfully to intermediate and long-run outcomes. Yet, reliable and consistently-reported measures of family environment paired with individual-level economic and social outcomes are rarely available over long time spans. Thus, we know very little about the contribution of the family environment to human capital formation in the US (and other countries) over the nineteenth and early-twentieth centuries and how it may have changed. This period spanned the demographic and health transitions (Greenwood and Seshadri 2002; Costa 2015) and the emergence of modern economic growth (Galor 2005). Individuals born in 1910 experienced rising life expectancy (a 30% increase at birth), education levels (a 35% increase in school enrollment rates and a 15-fold increase in high school graduation rates) and declining total fertility (by 40%) compared to the 1835 birth cohort. These trends were accompanied by institutional change – the abolition of slavery, compulsory schooling laws, and women's suffrage – and a historic macroeconomic shock in the form of the Great Depression. Furthermore, they may have interacted with the family environment in the production of human capital either as complements, thereby enhancing the firstborn advantage, or as substitutes, thereby diminishing within-family inequality.

In this paper, we use birth order as a within-family quasi experiment to study the effects of family environment on labor market and social outcomes for individuals born between 1835 and 1910. We ask the following research questions: How large were birth order effects on labor market outcomes? Were they stable or changing over time? Did birth order affect intergenerational occupational mobility? And finally, were the effects limited to economic outcomes or did birth order affect social outcomes as well? These questions are relevant to understanding the role that family environment played during the critical decades that shaped the modern US.

We answer these questions by leveraging a combination of large-scale census data and a sibling fixed effects estimation strategy. Our samples link historical US full-count census data in 30-year intervals: 1850-1880, 1870-1900, 1880-1910, 1900-1930, and 1910-1940.[1] While holding constant all time-invariant family characteristics, we compare outcomes for the firstborn (male) child in a family with a later-born brother.[2] In doing so, we build on a long history of studying birth order as a proxy for family circumstances in social psychology and economics (Adler 1928; Behrman and Taubman 1986; Black et al. 2005a; Roher et al. 2015; Breining et al. 2020).

We find remarkably stable birth order effects over the time frame considered. Our point estimate for the 1835-1850 birth cohorts implies a firstborn occupational premium of 2.5% of a standard deviation (SD) while the point estimate for the 1895-1910 birth cohorts is 2.6% of a SD. The three intermediate estimates range from 3.0% of a SD to 3.8% of a SD. We find similar patterns with alternative measures of occupational attainment such as the probability of working in a white collar occupation. Turning to intergenerational mobility, we find that firstborn sons were more likely to experience upward mobility and less likely to be downward mobile, suggesting that birth order could play a role in intergenerational mobility. The effects also extend beyond occupational premiums and into social outcomes. Firstborn males were less likely to remain single as adults, with effect sizes ranging from 10.4% to 15.8%, and more likely to have children, with effect sizes ranging from 2.4% to 3.3%. When we split the sample by race, the effects on labor market and social outcomes for White men are comparable to the overall findings. However, the coefficients on birth order for Black men are mixed signed and mostly statistically insignificant.

Our findings are robust. Point estimates and statistical significance are not materially affected by changes in the age ranges of children or mothers that we include in the sample, the choice of control group, adjusting for fertility stopping rules and selection, or the use of different time spans for linked census data.

This paper makes contributions to several literatures, most notably one that investigates the

[1] Since the 1890 US Census records were destroyed in a fire, we cannot create the 1860-1890 or 1890-1920 links. We also conduct analysis with 20-year and 40-year links (Online Appendix Tables A1 and A2).

[2] We limit the analysis to men due to differences in how censuses are linked for men and women.

role of family environment in human capital production through the lens of birth order. Prior work has documented stark differences between earlier and later born children in health (Jayachandran and Pande 2017; Brenøe and Molitor 2018; Abdellaoui et al. 2022), parental investments (Price 2008; Averett et al. 2011; Lehmann et al. 2018), test scores (Silles 2010; Haan et al. 2014), educational attainment (Black et al. 2005a; de Haan 2010; Abdellaoui et al. 2022; Clark and Cummins 2024), earnings (Black et al. 2005a), income (Abdellaoui et al. 2022), occupational attainment (Nuevo-Chiquero et al. 2023; Clark and Cummins 2024), farm inheritance (Haws et al. 2023), non-cognitive skills and personality (Roher et al. 2015; Black et al. 2018), and risky behaviors and criminal activity (Argys et al. 2006; Breining et al. 2020).[3] Within this literature, only Black et al. (2005a), Black et al. (2018), Abdellaoui et al. (2022), Nuevo-Chiquero et al. (2023), and Clark and Cummins (2024) document effects on labor market outcomes, while only Nuevo-Chiquero et al. (2023) investigate if birth order matters for intergenerational transmission between parents and their children.

Our contribution to this literature is twofold. First, we consider changes in birth order effects over time in the US during the late-nineteenth and early-twentieth centuries – finding that the firstborn premiums are remarkably stable over an extended period. And second, we document effects on both labor market and social outcomes which include marriage and fertility. In complementary ongoing work, both Nuevo-Chiquero et al. (2023) and Clark and Cummins (2024) investigate birth order effects in different historical contexts: the Netherlands and England, respectively. We differ from these two papers in several dimensions. Most obviously, we study the US rather than a European country. Given economic and social differences, it's unclear if the results of European studies would translate to the US setting. Indeed, some (but not all) of our results are consistent with findings in Nuevo-Chiquero et al. (2023), but they are meaningfully different from those reported by Clark and Cummins (2024). In addition, we highlight the evolution of the firstborn premium

[3] Most of these studies document that later-born children have worse outcomes. However, this result is not universal as Haan et al. (2014) find the opposite pattern in Ecuador where unlike in other countries, parents tend to invest more in later- rather than earlier-born children. This is especially pronounced in poor and low-educated families while in rich and high-educated families the effect reverses to a firstborn advantage. Given massive changes in development and poverty over the time period we study, it is plausible that firstborn premiums might have changed signs in the US over the course of our sample.

over time while these papers focus mostly on differences in the effects across birth order parities. Finally, along with Nuevo-Chiquero et al. (2023) we use full population censuses rather than a sample of families with rare surnames as in Clark and Cummins (2024).

Our results also speak to the literature on intergenerational inequality. Existing studies document correlations between parents and their children in education (Black et al. 2005b), preferences (Dohmen et al. 2012), health (Bütikofer et al. 2024), non-cognitive skills (Grönqvistt et al. 2017), socio-emotional skills (Attanasio et al. 2021), income (Chetty et al. 2014), and wealth (Black et al. 2020). Furthermore, Song et al. (2019) show that intergenerational occupational mobility in the US declined between 1850 and 2015 – with much of the decline occurring for the early birth cohorts that we consider in this paper. Ward (2023) re-examines these historical trends by accounting for measurement error and race, while Buckles et al. (2023b) additionally use high-recall links and study women. Their results suggest "low but increasing mobility" in contrast to Song et al. (2019) who find "high but decreasing mobility."[4] We contribute to this literature by documenting that firstborn premiums (intragenerational within-family inequality) influence the likelihood of the intergenerational transmission of occupations.

2 Data

This paper utilizes public-use versions of the 1850-1940 full-count decennial US Censuses of Population from IPUMS (Ruggles et al. 2023). The 1890 Census is omitted since the records were destroyed in a fire.[5] The full-count US decennial censuses include characteristics such as a person's name, birth year, birthplace, gender, race, place of residence, and occupation. The data are organized by household, allowing us to observe family relationships and characteristics of household members. We link individuals between the IPUMS versions of the full-count censuses using publicly available crosswalks from the Census Tree database (Price et al. 2021; Buckles et al. 2023a; Price et al. 2023).[6]

The starting point for studying the effects of birth order is identifying the set of individuals who

[4] In this paper we use the same Census Tree links as Buckles et al. (2023b).
[5] Enslaved Black individuals are not included in the 1850 sample.
[6] In Appendix Table A12 we show that our results are robust to alternative linking procedures.

were born to the same parents. In our baseline specifications we define siblings as those sharing a common father (or, if missing, a common mother), conditional on the parent(s) and children being enumerated in the same household in the census. In doing so, we exclude stepfathers from the pool of parents due to potential ambiguity in shared parentage. We define birth order among siblings based on relative ages and the order of enumeration within a household (those listed first are considered older when more than one child in a sibship reports the same age).

Next, we impose a series of household-level sample restrictions that address measurement error in birth order and lead to our estimation sample.[7] First, we drop households in which the oldest sibling is above the age of 15. As the share of children living with their parents tended to decline during the teenage years, the presence of a sibling older than age 15 increases the likelihood that an even older sibling had formed their own household and is not observed together with their parents.[8] In such cases, the risk of misclassifying birth order is elevated. Next, we drop sibling groups that include foreign-born children since the potential presence of siblings abroad is another source of measurement error in birth order. Then, we drop mixed-race sibling groups and restrict to individuals who report their race as Black or White in order to avoid conflating birth order differences with unobserved racial differences.[9] Finally, we subset to males only and households in which the firstborn child is a male due to gender-specific differences in the census linking procedure.[10]

Since census enumerators did not ask about biological parents, another potential source of measurement error in sibling status and birth order stems from the fact that we identify siblings based on grouping individuals by common social environment and guardianship (i.e. nurture not nature). For instance, given that familial ties are defined in relation to the household head, who

[7] Prior to the steps discussed below, we drop a small number of observations with duplicate HISTIDs, the IPUMS unique identifier, for which the uniqueness of an individual record is unclear. We also drop a handful of outlier households in which the number of siblings exceeds 15.

[8] We deal with this issue by further restricting our analysis to families in which the oldest child in the household is age 12 or younger (Panels A and B of Table A3) and find that our main results barely change.

[9] The granularity of the racial categories varies across censuses. In the 1900-1920 censuses, some mixed-race individuals report their race as "mulatto." We exclude these individuals from the estimation sample to maintain a consistent and comparable classification of race across all censuses.

[10] While the Census Tree dataset includes a large number of women, these links are predominantly based on pre-existing links in FamilySearch's Family Tree which were created by users of the website and may not be representative of the population. Due to name changes at marriage, it is not possible to reliably link women using the machine learning model XGBoost (Buckles et al. 2023b).

was typically a man, some siblings may not share the same mother. Relatedly, our birth order measure may not align with biological birth order when children did not survive long enough to be enumerated in the census. Since we posit that the effects of birth order have a social or familial origin rather than a biological origin, and are thus interested in the birth order of surviving children, we are unconcerned by the deviation in these measures of birth order due to children who die in infancy or very early childhood.

Our primary outcome variable is an occupational income score derived from the occupational title reported in the census when an individual is ages 30-45 (Ruggles et al. 2023). We use the OCCSCORE variable from IPUMS and standardize it across the full population by gender within each census year.[11] The use of OCCSCORE facilitates comparisons over time while other socioeconomic measures such as income, literacy, or education are not consistently reported across censuses. We also construct binary measures of occupational status based on the OCC1950 variable from IPUMS that follows the 1950 Census Bureau classification with 283 occupational categories.[12] Lastly, we construct social outcomes using information in the census when individuals are ages 30-45. We define an indicator for being single based on reported marital status and measures of fertility based on the presence or number of children enumerated in the same household.[13] After imposing the sample restrictions, our estimation sample includes firstborn-male sibships in which the firstborn and at least one younger male sibling are linked to the census during adulthood. For labor market outcomes, we also require individuals to report a valid occupational income score.

[11] OCCSCORE "assigns each occupation in all years a value representing the median total income (in hundreds of 1950 dollars) of all persons with that particular occupation in 1950." Zero represents missing (Ruggles et al. 2023).

[12] We use OCC1950 codes to assign an occupation to one of four categories: white collar, skilled, farmers, and unskilled. White-collar occupations include professionals (OCC1950 codes 0-99), managers (200-299), clerical workers (300-399), and sales workers (400-499). Skilled workers are craftsmen (500-599). Farmers include farm owners and tenants (OCC1950 code 100) in addition to farm managers (OCC1950 code 123). Unskilled workers include operatives (600-699), service workers (700-799), farm laborers and general laborers (800-970). Occupations with codes above 970 are excluded from the sample as they signify being out of the labor force.

[13] We code an individual as single if their reported marital status is "Never married/single." Other response categories are: married (with or without spouse present in the household), separated, divorced, or widowed.

3 Methods

The main estimating equation compares a firstborn boy to his brother who is closest in birth order and linked across censuses:

$$Y_{ijcms} = \beta_0 + \beta_1 \text{firstborn}_i + \mathbf{\Theta X_i} + \alpha_j + \gamma_c + \phi_m + \delta_s + \varepsilon_{ijcms} \tag{1}$$

where i denotes individuals, j denotes families, c denotes birth cohort, m denotes mother's age at birth, and s denotes state of birth. A vector of control variables, X_i, that is used in select specifications includes a race indicator as well as fixed effects for family size and father's occupational status.[14] All regressions control for year of birth (γ_c) fixed effects. Our preferred specification further includes a family fixed effect, α_j, which effectively forces the comparison of outcomes to be within a family, and fixed effects that vary at the individual-level for mother's age at birth (ϕ_m) and state of birth (δ_s). The family fixed effects are appealing because they control for factors that are invariant within families and might be unobserved or observed with measurement error in historical records. Thus, in this equation, β_1 compares outcomes of the firstborn son in the family with his younger brother closest in age and linked to the census in adulthood. We cluster standard errors, ε_{ijcms}, at the family level.

There are two issues related to estimating birth order effects that need to be addressed. First, family size can have its own effects on children other than directly through birth order. For example, Bagger et al. (2021) note that birth order and family size are jointly determined, i.e., one cannot manipulate family size while holding the within-family distribution of birth order constant. Second, families with same-sex children among their firstborns are more likely to have additional births — a feature that is often used as an instrument for family size (Angrist et al. 2010; Conley and Glauber 2006; Jones et al. 2023).[15] We take four steps to deal with family size and sex composition. First, in specifications that exclude family fixed effects, we show that the firstborn premium

[14] We use the same categorization procedure for fathers occupations as for occupational outcomes of sons.

[15] Black et al. (2005a) also use this instrument in their analysis of family size and note that same-sex siblings could have an independent effect on child outcomes and that the birth order effects are both stable when instrumenting for completed family size and orders of magnitude larger than the family size effects. See Tan (2019) for a discussion of family size effects in a historical context.

is larger without controls for family size. Second, in our preferred specifications, we include family fixed effects which control for family size and its potential effects on the outcomes. Third, we report results for different sex compositions including families in which the firstborn child is a girl (Table A4). Fourth, we show estimates separately for smaller and larger families (Table A5).

Another issue is that if a child requires costly parental investments, for example due to health problems (Ejrnaes and Pörtner 2004), parents may choose not to continue having children and thus last-born children may be negatively selected, thereby biasing the firstborn premium upward. We address this issue by excluding families in which the firstborn child is compared to the youngest male child enumerated in the household (Panel E of Table A3).

Finally, a more nuanced point relates to the interpretation of our estimates in the context of nineteenth and early-twentieth century inheritance laws and norms. Specifically, if primogeniture – a historical practice in which the eldest son inherited a family's land and titles – was common during our study period, then the firstborn advantage could solely reflect prevailing laws and customs rather than effects of family environment. Shammas et al. (1987) highlights (in their Table 3.1) that by 1790 most colonies had equal rights to inheritance and that "by 1800 in most states, sons and daughters received equal shares in real and personal property." This suggest that although primogeniture was initially in place in the US, it was no longer a common practice by the time the earliest cohorts in our sample were born in 1835. Furthermore, Haws et al. (2023) find a limited role of birth order in farm inheritances in the US. While their 1900-1920 estimates imply a small birth order penalty, it vanishes in census links over 30 to 40 years. Based on these facts it is unlikely that inheritance customs played a meaningful role in our setting and thus we prefer to interpret the estimates as the effects of family environment.

4 Results

Table 1 presents our main results for long-run occupational outcomes measured at ages 30-45. Each coefficient comes from a separate regression. Columns 1 to 5 present results using different linked samples while column 6 pools all observations. Panels A and B compare first- and later-born boys across families while Panels C and D restrict to within-family comparisons. The inclusion of control variables in Panel B reduces the magnitude of the firstborn premium by 23-51%. The addition of family size fixed effects accounts for much of this difference. The firstborn premium is slightly larger when we make within-family comparisons in Panel C, while adding within-family controls in Panel D does not meaningfully alter the estimates. Our preferred estimates in Panel D imply 2.5-3.8% of a SD higher occupational income scores for the firstborn compared with later-born boys.[16] The effect size is 2.5% of a SD for the 1835-1850 birth cohorts, 2.6% of a SD for individuals born 1895-1910, and up to 3.8% for cohorts in between. When we pool together all census waves we estimate a firstborn advantage of 3.2% of a SD for children born between 1835 and 1910.[17] Overall, birth order effects appear to be very stable over an extended period of time.

In Table 2 we further investigate whether the occupational effects of birth order are concentrated in specific parts of the distribution. We group occupations into four categories based on OCC1950 codes (see Section 3 for details). Panels A and B focus on lower-status occupations. In Panel A, the dependent variable is an indicator that equals one if an individual worked as a farmer (including farm owners, managers, and tenants). In Panel B, it is an indicator for unskilled occupations in non-farm sectors. In both cases, firstborns are less likely to work in these occupations, with larger effect sizes for farming in earlier census waves and increasing magnitudes over time for unskilled occupations. This pattern is consistent with the changing occupational structure of the economy as farming was dominant in the nineteenth century while unskilled employment grew in other sectors in the early twentieth century. Panels C and D focus on the higher-status skilled and white collar occupations. Firstborns are more likely to enter these occupations and do so at a

[16] At the beginning of our sample period, this occupational upgrading relative to the mean could represent moving from a general laborer to an apprentice in a trade, and at the end this would be consistent with moving from a sales clerk to a general clerical worker.

[17] In Table A1 we verify that this pattern is not an artifact of using the 30-years links. We find a very similar pattern of estimates for the firstborn premium when using 20-year links, which include an additional census wave, with slightly smaller effects for the earliest (1850-1870) and latest (1920-1940) census waves at 3.2% and 3.8% of a SD, respectively, compared to 3.5-5.0% of a SD for intervening cohorts. We do not include the results based on the 20-year links among our preferred estimates because they measure occupational outcomes at ages 20-35 (rather than ages 30-45) which might be an imperfect proxy for life-cycle occupation – especially for the more recent cohorts for whom educational attainment increased substantially. This concern is particularly relevant for birth order as the later-born child is more likely to be on the steeper part of the curve for life-cycle labor market outcomes. In Table A2 we report estimates using 40-years links in which occupational outcomes are measured at ages 40-55. Again, we find stable firstborn premiums starting with an estimate of 3.9% of a SD for the 1860-1900 link and ending with an estimate of 3.1% of a SD for the 1900-1940 link.

consistent rate across the time span of our analysis (5.1-6.4% for white collar occupations and 2.9-4.2% for white collar or skilled occupations). Thus, it appears that firstborn premiums affect the entire distribution of occupational outcomes by keeping individuals out of relatively lower-skilled jobs and enabling them to enter higher-status occupations.

The magnitudes of our firstborn premium estimates are generally similar to or smaller than findings from other historical and modern contexts. To facilitate a comparison with Nuevo-Chiquero et al. (2023), we compute the "adjusted Song score" as an outcome that can be interpreted as an occupational education percentile score (Song et al. 2019; Ward 2023). Using a sample that pools all linked census samples, we find a firstborn premium of 1.014 percentile points (Panel D of Column 1 in Table A7), an effect size that is very similar to the second-born penalty of approximately 1 percentile point found by Nuevo-Chiquero et al. (2023) for the Netherlands.[18] Likewise, Clark and Cummins (2024) present their occupational outcomes on a 0-100 scale which is similar to the percentile measure. They find no effects for lineages that are representative of the UK population, with a coefficient of 0.239 that is less than a quarter of the size of our preferred estimate. On the other hand, their statistically significant estimate for the very wealthy is 1.785 which is almost double the size of our coefficient estimate of 0.974 for sons with white collar fathers. In addition, our estimates appear smaller than the findings of two studies on birth order and labor market outcomes using modern data. Black et al. (2005a) use Norwegian administrative data on earnings for 1986-2000 and find a firstborn premium of 2-3%. Black et al. (2018) use Swedish administrative data on occupations for 1996-2000 and find a second-born penalty of approximately 15% for managers whose occupational rank we view as most comparable to the white-collar occupations in our data.

Next, we show that birth order moderates intergenerational occupational mobility. We follow the approach used by Song et al. (2019) and define immobility as the likelihood that a son stayed in the same micro-class occupation as their father.[19] The results of this analysis are presented

[18]We hesitate to compare the effects of the firstborn premium on the adjusted Song score across linked census waves for two reasons. First, the data inputs used to compute the score differ across census waves: the 1850-1930 censuses report literacy while the 1940 census reports years of education. Second, the composition of ages at which the outcomes used to compute the adjusted Song score are observed differs across birth cohorts because the 1890 census is missing.

[19]The micro-class occupational scheme aggregates the 283 occupational categories of the IPUMS OCC1950 vari-

in Panel A of Table 3. The sample immobility rates declined from 42.5% for the 1850-1880 linked sample to 21.1% for the 1910-1940 linked sample, which is consistent with Song et al. (2019). Our results suggest that although birth order effects are not statistically significant for the early cohorts, firstborns among the later cohorts are less likely to be immobile. The effect sizes range from 2.8% for the 1880-1910 linked sample to 1.9% for the 1910-1940 linked sample. In contrast, Nuevo-Chiquero et al. (2023) find that firstborns in the Netherlands are more likely to end up in the same occupation as their father. We then compare the OCCSCORE of fathers and sons to examine effects on absolute upward (Panel B) and downward (Panel C) mobility. We find statistically significant increases in the probability of upward mobility ranging from 1.5% to 3.8% and decreases in the probability of downward mobility between 1.8% and 3.1% across all linkage pairs – a clear indication of a firstborn advantage that nevertheless appears to be smaller than the approximately 10% effect of lower health capital in childhood on absolute mobility that has been estimated for a similar time period in the UK (Karbownik and Wray 2024).

Finally, we examine effects of birth order on social outcomes. Unlike prior research, we cannot consistently observe educational (de Haan 2010; Haan et al. 2014) or behavioral and non-cognitive outcomes (Black et al. 2018; Breining et al. 2020). Instead, in Table 4 we study three social outcomes that we can observe for all adult males at ages 30-45: the probability of being single (Panel A), the probability of having any children (Panel B), and the number of children conditional on having any (Panel C). Effects on fertility are of interest given that the fraction of people having children declined by 16% and the average number of children per family fell by 26% across the 1835-1910 birth cohorts. Despite these changes we find very stable results: firstborns are less likely to remain single, they are more likely to have children, and they have more children conditional on having any. As with the labor market outcomes, the effect sizes are stable over time at 10.4%-15.8%, 2.4%-3.3%, and 2.1%-3.4% for the three outcomes, respectively. Especially for fertility, the variability in effect sizes appears tiny compared with the cross-cohort changes in the outcomes.

We stratify Black and White families to investigate the effects of birth order separately by race.

able to 70 groups (Song et al. 2019).

We report results for occupational and social outcomes of Black and White men in Tables A8 and A9, respectively. Our results are driven by White families as we do not find any evidence of a firstborn premium in occupational or social outcomes for Black families. The firstborn premium is consistently positive and statistically significant at conventional levels only for White individuals, while for Black individuals we mostly find insignificant and mixed-signed coefficients.[20]

Lastly, we explore heterogeneity along four dimensions: birth parity (Table A10), family size (Table A5), mother's age at first birth (Table A6), and father's socioeconomic status (Table A11). We find larger firstborn premiums when comparisons are made to younger siblings further apart in birth order, which suggests that our main estimates may understate the disadvantage of later-born siblings. When we split the sample by family size, we find modestly larger effects for above-median sized families, although this difference vanishes for the 1910-1940 linkage. On the other hand, effects are similar across households with mothers whose age at first birth was above versus below the median age, with the exception of the 1850-1880 sample. Finally, the estimates for earlier cohorts are somewhat larger for lower socioeconomic status families, but the magnitude of this difference depends on the exact census pair considered.

5 Robustness Checks

This section presents a series of analyses to gauge the robustness of our main results. We have already documented in Tables A1 and A2 that our results are unaffected by changing the gap between the baseline and follow-up years when linking census waves. Therefore, for brevity, we focus on the 30-year links and family fixed effects models (Panel D of Table 1).

We first address the possibility that birth order is assigned incorrectly when older siblings have formed their own households, and therefore are not enumerated in the census with their parents and younger siblings. Panel A of Table A3 restricts the maximum age of a child in the household to be 12 years in the baseline census (rather than 15 years) while Panel B further restricts the sample to

[20]In the three largest samples – 1880-1910, 1900-1930, and 1910-1940 – we can reject that birth order effects on occupational income score for Whites and Blacks are statistically identical. One hypothesis explaining this discrepancy could be the role of widespread labor market discrimination that may have limited firstborn African-American men's ability to convert higher human capital into better occupations.

younger mothers who were between the ages 18-30 when any of their children were born. Despite a reduction in the sample size by half, the magnitude of our estimates and their pattern over time remain very similar to our preferred specification.

Next we consider different comparison groups of siblings. Recall that Table 1 compares the firstborn (male) child to the younger brother closest in birth order who can be linked to outcomes in adulthood, ignoring siblings born in between. Our conclusions remain unchanged when we compare the firstborn (male) child to all later-born brothers linked to outcomes (Panel C of Table A3). Then in Table A4 we estimate the second-born vs. next-born boy effects for families in which the firstborn child is a girl. We estimate statistically insignificant and negative signed coefficients for the first two census waves, but for subsequent linkages we find positive and statistically significant second-born premiums that are 58-74% smaller than estimates in Panel D of Table 1. This result is consistent with prior literature that finds a disproportionate premium for being a firstborn that declines in birth order (Black et al. 2005a).

Returning to Table A3, the final two panels tackle the issue of endogenous fertility stopping. We exclude families with exactly two children (Panel D) or families for which the younger sibling included in the estimation sample is the youngest in the household (Panel E). These additional sample restrictions do not affect our results.

We also explore the sensitivity of our results to the choice of linking method and concerns about the representativeness of our linked samples. Our results remain unchanged when we restrict to the subset of links in the Census Tree dataset that are included in the Census Linking Project crosswalks (Abramitzky et al. 2020), are generated by the XGBoost machine learning algorithm, or appear in the Family Tree (Appendix Table A12). We also implement a version of the Bailey et al. (2020) procedure that weights the households in our estimation samples so that they are representative of the baseline population of interest in the childhood census (Appendix Table A13). The magnitude and stability of the firstborn premiums remain unchanged. Overall, we conclude that our results are robust to a reasonable set of alternative sample and estimation choices.

6 Conclusions

Birth order affects a multitude of short- and long-run outcomes in contemporary data spanning developed and developing countries. Despite its importance for economic outcomes and human capital development, there is limited evidence if and to what extent birth order mattered historically, when fertility and mortality rates were much higher and the quality of parental inputs and knowledge about returns to investment in children were much lower. We contribute to this literature by estimating firstborn premiums for US-born individuals between 1835 and 1910, paying particular attention to the dynamics of the estimates over time.

We find evidence that the firstborn premium was present among families as early as those raising their children in the 1830s. Remarkably, the effects identified for these cohorts are very similar to those who grew up in the first decades of the twentieth century. They extend beyond occupational outcomes and into the social domain including the likelihood of remaining single as well as fertility at both the extensive and intensive margins. Finally, we find that birth order affected the intergenerational transmission of occupations between fathers and sons. Irrespective of the outcome and the period in US history, we find that firstborn males were always better off compared to their younger brothers.

Our results have important implications for understanding the role that family resources play in shaping human capital and labor market outcomes of children. To the extent that birth order credibly estimates the marginal returns to family environment, our findings suggest that these returns changed very little over an extended period of time during which the American family experienced dramatic cultural, economic, and social changes.

References

Abdellaoui, Abdel, Oana Borcan, Pierre-Ander Chiappori, and David Hugh-Jones (2022) "Trading social status for genetics in marriage markets: Evidence from UK Biobank."

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, Santiago Pérez, and Myera Rashid (2020) "Census Linking Project: Version 2.0 [dataset].," https://censuslinkingproject.org.

Adler, Alfred (1928) "Characteristics of the first, second, and third children," *Children*, Vol. 3, No. 5, p. 14.

Almond, Douglas and Janet Currie (2011) "Human capital development before age five," in David Card and Orley Ashenfelter eds. *Handbook of Labor Economics*, Vol. 4B, pp. 1315–1486.

Almond, Douglas, Janet Currie, and Valentina Duque (2018) "Childhood circumstances and adult outcomes: Act II," *Journal of Economic Literature*, Vol. 56, No. 4, pp. 1360–1446.

Angrist, Joshua, Victor Lavy, and Analia Schlosser (2010) "Multiple experiments for the causal link between the quantity and quality of children," *Journal of Labor Economics*, Vol. 28, No. 4, pp. 773–824.

Argys, Laura, Daniel Rees, Susan Averett, and Benjama Witoonchart (2006) "Birth order and risky adolescent behavior," *Economic Inquiry*, Vol. 44, No. 2, pp. 215–233.

Attanasio, Orazio, Aureo de Paula, and Alessandro Toppeta (2021) "The persistence of socio-emotional skills: Life cycle and intergenerational evidence," Working Paper 27823, National Bureau of Economic Research.

Averett, Susan, Laura Argys, and Daniel Rees (2011) "Older siblings and adolescent risky behavior: Does parenting play a role?," *Journal of Population Economics*, Vol. 24, No. 3, pp. 957–978.

Bagger, Jesper, Javier Birchenall, Hani Mansour, and Sergio Urzu'a (2021) "Education, birth order and family size," *Economic Journal*, Vol. 131, No. 633, pp. 33–69.

Bailey, Martha, Connor Cole, and Catherine Massey (2020) "Simple Strategies for Improving Inference with Linked Data: A Case Study of the 1850-1930 IPUMS Linked Representative Historical Samples," *Historical Methods*, Vol. 53, No. 2, pp. 80–93.

Behrman, Jere and Paul Taubman (1986) "Birth order, schooling, and earnings," *Journal of Labor Economics*, Vol. 4, No. 3, pp. 121–145.

Black, Sandra, Paul Devereux, Petter Lundborg, and Kaveh Majlesi (2020) "Poor little rich kids? The role of nature versus nurture in wealth and other economic outcomes and behaviors," *Review of Economic Studies*, Vol. 87, No. 4, pp. 1683–1725.

Black, Sandra, Paul Devereux, and Kjell Salvanes (2005a) "The more the merrier? The effects of family size and birth order on children's education," *Quarterly Journal of Economics*, Vol. 120, No. 2, pp. 669–700.

——— (2005b) "Why the apple doesn't fall far: Understanding intergenerational transmission of human capita," *American Economic Review*, Vol. 95, No. 1, pp. 437–449.

Black, Sandra, Erik Grönqvist, and Öckert Björn (2018) "Born to lead? The effect of birth order on non-cognitive skills," *Review of Economics and Statistics*, Vol. 100, No. 2, pp. 274–286.

Breining, Sanni, Joseph Doyle, David Figlio, Krzysztof Karbownik, and Jeffrey Roth (2020) "Birth order and delinquency: Evidence from Denmark and Florida," *Journal of Labor Economics*, Vol. 38, No. 1, pp. 95–142.

Brenøe, Anne and Ramona Molitor (2018) "Birth order and health of newborns: What can we learn from Danish registry data?," *Journal of Population Economics*, Vol. 31, pp. 363–395.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley E.B. Wilbert (2023a) "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project," NBER Working Paper 31671.

Buckles, Kasey, Joseph Price, Zachary Ward, and Haley Wilbert (2023b) "Family Trees and Falling Apples: Historical Intergenerational Mobility Estimates for Women and Men," Working Paper.

Bütikofer, Aline, Rita Ginja, Krzysztof Karbownik, and Fanny Landaud (2024) "(Breaking) intergenerational transmission of mental health," *Journal of Human Resources*, Vol. forthcoming.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014) "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *Quarterly Journal of Economics*, Vol. 129, No. 4, pp. 1553–1623.

Clark, Gregory and Neil Cummins (2024) "Birth order and social outcomes, England, 1680-2024."

Conley, Dalton and Rebecca Glauber (2006) "Parental educational investment and children's academic risk: Estimates of the impact of sibship size and birth order from exogenous variation in fertility," *Journal of Human Resources*, Vol. 41, No. 4, pp. 722–737.

Costa, Dora (2015) "Health and the economy in the United States from 1750 to the present," *Journal of Economic Literature*, Vol. 53, No. 3, pp. 503–570.

Cunha, Flavio, James Heckman, and Susanne Schennach (2010) "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, Vol. 78, No. 3, pp. 883–931.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2012) "The intergenerational transmission of risk and trust attitudes," *Review of Economic Studies*, Vol. 79, No. 2, pp. 645–677.

Ejrnaes, Mette and Claus Pörtner (2004) "Birth order and the intrahousehold allocation of time and education," *Review of Economics and Statistics*, Vol. 86, No. 4, pp. 1008–1019.

Galor, Oded (2005) "Chapter 4 From Stagnation to Growth: Unified Growth Theory," in Philippe Aghion and Steven N. Durlauf eds. *Handbook of Economic Growth*, Vol. 1: Elsevier, pp. 171–293.

Greenwood, Jeremy and Ananth Seshadri (2002) “The US demographic transition,” *American Economic Review*, Vol. 92, No. 2, pp. 153–159.

Grönqvistt, Erik, Björn Öckert, and Jonas Vlachos (2017) “The intergenerational transmission of cognitive and noncognitive abilities,” *Journal of Human Resources*, Vol. 52, No. 4, pp. 887–918.

de Haan, Monique (2010) “Birth order, family size and educational attainment,” *Economics of Education Review*, Vol. 29, No. 4, pp. 576–588.

de Haan, Monique, Erik Plug, and Jose Rosero (2014) “Birth order and human capital development: Evidence from Ecuador,” *Journal of Human Resources*, Vol. 49, No. 2, pp. 359–392.

Haws, Adrian, David Just, and Joseph Price (2023) “Who (actually) gets the farm? Intergenerational farm succession in the United States,” NBER Working Paper 31604.

Jayachandran, Seema and Rohini Pande (2017) “Why are Indian children so short? The role of birth order and son preference,” *American Economic Review*, Vol. 107, No. 9, pp. 2600–2629.

Jones, Todd, Matthew Millington, and Joseph Price (2023) “Changes in parental gender preference in the USA: Evidence from 1850 to 2019,” *Journal of Population Economics*, Vol. 36, No. 4, pp. 3057–3070.

Karbownik, Krzysztof and Anthony Wray (2024) “Lifetime and intergenerational consequences of poor childhood health,” *Journal of Human Resources*, Vol. forthcoming.

Lehmann, Jee-Yeon, Ana Nuevo-Chiquero, and Marian Vidal-Fernandez (2018) “The early origins of birth order differences in children’s outcomes and parental behavior,” *Journal of Human Resources*, Vol. 53, No. 1, pp. 123–156.

Nuevo-Chiquero, Ana, Marian Vidal-Fernandez, and Jee-Yeon Lehmann (2023) “The birth order effect: A modern phenomenon.”

Price, Joseph (2008) “Parent-child quality time: Does birth order matter?,” *Journal of Human Resources*, Vol. 43, No. 1, pp. 240–265.

Price, Joseph, Kasey Buckles, Adrian Haws, , and Haley Wilbert (2023) *The Census Tree [dataset]*, Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

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*, Vol. 80, p. 101391.

Roher, Julia, Boris Egloff, and Stefan Schmukle (2015) “Examining the effects of birth order on personality,” *Proceedings of the National Academy of Sciences*, Vol. 112, No. 46, pp. 14224–14229.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Anne Chen, Grace Cooper, Stephanie Richards, and Megan Schouweiler (2023) *IPUMS USA: Version 14.0 [dataset]*, Minneapolis, MN: IPUMS.

Shammas, Carole, Marylynn Salomon, and Michel Dahlin (1987) *Inheritance in America: From colonial times to the present*, New Brunswick and London: Rutgers University Press.

Silles, Mary (2010) “The implications of family size and birth order for test scores and behavioral development,” *Economic of Education Review*, Vol. 29, No. 5, pp. 795–803.

Song, Xi, Catherine Massey, Karen Rolf, Joseph Ferrie, Jonathan Rothbaum, and Yu Xie (2019) “Long-term decline in intergenerational mobility in the United States since the 1850s,” *Proceedings of the National Academy of Sciences*, Vol. 117, No. 1, pp. 251–258.

Tan, Hui Ren (2019) “More is less? The impact of family size on education outcomes in the United States, 1850-1940,” *Journal of Human Resources*, Vol. 54, No. 4, pp. 1154–1181.

Ward, Zachary (2023) “Intergenerational Mobility in American History: Accounting for Race and Measurement Error,” *American Economic Review*, Vol. 113, No. 12, pp. 3213–48.

7 Tables

Table 1. Firstborn Premium for Occupation Status of Men

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. OLS. No controls except birth year fixed effects.						
Firstborn (=1)	0.041	0.031	0.048	0.044	0.032	0.039
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
Mean of Y	0.189	0.173	0.196	0.202	0.183	0.190
# of Households	164,142	295,258	474,942	722,628	874,222	2,531,192
Observations	328,284	590,516	949,884	1,445,256	1,748,444	5,062,384
B. OLS. Family and individual controls.						
Firstborn (=1)	0.020	0.024	0.033	0.026	0.018	0.024
	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
Mean of Y	0.189	0.173	0.196	0.202	0.183	0.190
# of Households	164,142	295,258	474,942	722,628	874,222	2,531,192
Observations	328,284	590,516	949,884	1,445,256	1,748,444	5,062,384
C. Family fixed effects. No controls except birth year fixed effects.						
Firstborn (=1)	0.026	0.027	0.037	0.038	0.024	0.031
	(0.006)	(0.005)	(0.004)	(0.003)	(0.002)	(0.001)
Mean of Y	0.189	0.173	0.196	0.202	0.183	0.190
# of Households	164,142	295,258	474,942	722,628	874,222	2,531,192
Observations	328,284	590,516	949,884	1,445,256	1,748,444	5,062,384
D. Family fixed effects. Within-family controls.						
Firstborn (=1)	0.025	0.030	0.037	0.038	0.026	0.032
	(0.007)	(0.005)	(0.004)	(0.003)	(0.002)	(0.001)
Mean of Y	0.189	0.173	0.196	0.202	0.183	0.190
# of Households	164,142	295,258	474,942	722,628	874,222	2,531,192
Observations	328,284	590,516	949,884	1,445,256	1,748,444	5,062,384

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include year of birth fixed effects. Panel B controls for race and adds fixed effects at the household level for family size and father's occupational status in addition to individual-level fixed effects for state of birth and mother's age at birth. We use OCC1950 codes to assign the father's occupation to one of four categories: white collar, skilled, farmers, and unskilled. See Section 3 for details. Panels C and D include family FEs. Panel D includes state of birth and mother's age at birth FEs. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table 2. Firstborn Premium for Additional Occupational Outcomes of Men

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. P(Farmer)						
Firstborn (=1)	–0.009	–0.004	–0.011	–0.005	–0.002	–0.005
	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Mean of Y	0.459	0.359	0.315	0.205	0.145	0.240
# of Households	164,142	295,258	474,942	722,628	874,222	2,531,192
Observations	328,284	590,516	949,884	1,445,256	1,748,444	5,062,384
B. P(Unskilled)						
Firstborn (=1)	–0.000	–0.009	–0.007	–0.012	–0.014	–0.011
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Mean of Y	0.235	0.291	0.277	0.297	0.368	0.313
# of Households	164,142	295,258	474,942	722,628	874,222	2,531,192
Observations	328,284	590,516	949,884	1,445,256	1,748,444	5,062,384
C. P(White collar)						
Firstborn (=1)	0.009	0.013	0.015	0.019	0.020	0.017
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Mean of Y	0.178	0.204	0.249	0.312	0.316	0.280
# of Households	164,142	295,258	474,942	722,628	874,222	2,531,192
Observations	328,284	590,516	949,884	1,445,256	1,748,444	5,062,384
D. P(Skilled or white collar)						
Firstborn (=1)	0.009	0.013	0.017	0.017	0.016	0.016
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Mean of Y	0.306	0.349	0.407	0.498	0.487	0.447
# of Households	164,142	295,258	474,942	722,628	874,222	2,531,192
Observations	328,284	590,516	949,884	1,445,256	1,748,444	5,062,384

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable in each panel is an indicator equal to one if an individual is employed in the stated occupational category. We use OCC1950 codes to assign an individual's occupation to one of four categories: white collar, skilled, farmers, and unskilled. White-collar occupations include professionals (OCC1950 codes 0-99), managers (200-299), clerical workers (300-399), and sales workers (400-499). Skilled workers are craftsmen (500-599). Farmers include farm owners and tenants (OCC1950 code 100) in addition to farm managers (OCC1950 code 123). Unskilled workers include operatives (600-699), service workers (700-799), farm laborers and general laborers (800-970). The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects – corresponding to the specification in Panel D of Table 1. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table 3. Absolute Mobility

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. P(Same microclass occupation as father)						
Firstborn (=1)	–0.005	–0.002	–0.009	–0.006	–0.004	–0.006
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Mean of Y	0.425	0.347	0.327	0.258	0.211	0.277
# of Households	153,347	258,668	445,648	627,770	760,229	2,245,662
Observations	306,694	517,336	891,296	1,255,540	1,520,458	4,491,324
B. P(Upward mobility)						
Firstborn (=1)	0.012	0.009	0.015	0.012	0.008	0.011
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Mean of Y	0.319	0.431	0.452	0.523	0.528	0.486
# of Households	153,347	258,693	445,671	627,783	760,261	2,245,755
Observations	306,694	517,386	891,342	1,255,566	1,520,522	4,491,510
C. P(Downward mobility)						
Firstborn (=1)	–0.005	–0.006	–0.006	–0.007	–0.005	–0.006
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Mean of Y	0.262	0.226	0.228	0.227	0.273	0.245
# of Households	153,347	258,693	445,671	627,783	760,261	2,245,755
Observations	306,694	517,386	891,342	1,255,566	1,520,522	4,491,510

Notes: Each coefficient comes from a separate regression on male-only samples. In Panel A the dependent variable is an indicator equal to one if an individual has the same micro-class occupation as his father (Song et al. 2019). In Panels B and C the dependent variables are indicators for whether the OCCSCORE (Ruggles et al. 2023) of an individual's occupation is higher or lower than his father's OCCSCORE, respectively. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects – corresponding to the specification in Panel D of Table 1. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table 4. Firstborn Premium for Social Outcomes of Men

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. P(Single)						
Firstborn (=1)	–0.018	–0.019	–0.021	–0.022	–0.020	–0.021
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Mean of Y	0.132	0.182	0.166	0.139	0.148	0.152
# of Households	173,078	354,025	519,221	793,631	931,865	2,771,820
Observations	346,156	708,050	1,038,442	1,587,262	1,863,730	5,543,640
B. P(Any children)						
Firstborn (=1)	0.019	0.016	0.019	0.021	0.021	0.020
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Mean of Y	0.758	0.675	0.672	0.672	0.638	0.666
# of Households	173,078	354,008	518,871	793,629	931,865	2,771,451
Observations	346,156	708,016	1,037,742	1,587,258	1,863,730	5,542,902
C. Number of children — any kids						
Firstborn (=1)	0.116	0.086	0.075	0.075	0.055	0.073
	(0.014)	(0.011)	(0.009)	(0.007)	(0.006)	(0.004)
Mean of Y	3.462	3.262	3.147	2.820	2.574	2.913
# of Households	108,543	183,803	268,902	406,984	436,581	1,404,813
Observations	217,086	367,606	537,804	813,968	873,162	2,809,626

Notes: Each coefficient comes from a separate regression on male-only samples. P(Single) is an indicator variable for reporting marital status as "never married/single" in the census (Panel A). P(Any children) is an indicator variable for having at least one child in the target census (Panel B). Number of children is count of children in the target census conditional on having any children (Panel C). The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects – corresponding to the specification in Panel D of Table 1. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Appendix: For Online Publication

Birth order in the very long-run: Estimating firstborn premiums between 1850 and 1940

Angela Cools, Jared Grooms, Krzysztof Karbownik,
Siobhan O'Keefe, Joseph Price, and Anthony Wray

A Online Appendix Tables

Table A1. Robustness of Firstborn Premium to 20-year gap

	(1) 1850 to 1870	(2) 1860 to 1880	(3) 1880 to 1900	(4) 1900 to 1920	(5) 1910 to 1930	(6) 1920 to 1940	(7) Pooled
A. OLS. No controls except birth year fixed effects.							
Firstborn (=1)	0.035	0.060	0.034	0.038	0.019	0.019	0.029
	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Mean of Y	0.039	0.048	0.021	0.023	0.016	–0.017	0.012
# of Households	144,991	269,384	473,647	717,637	868,932	1,084,510	3,559,101
Observations	289,982	538,768	947,294	1,435,274	1,737,864	2,169,020	7,118,202
B. OLS. Family and individual controls.							
Firstborn (=1)	0.013	0.020	0.020	0.018	0.009	0.004	0.011
	(0.005)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Mean of Y	0.039	0.048	0.021	0.023	0.016	–0.017	0.012
# of Households	144,991	269,384	473,647	717,637	868,932	1,084,510	3,559,101
Observations	289,980	538,764	947,294	1,435,274	1,737,864	2,169,020	7,118,196
C. Family fixed effects. No controls except birth year fixed effects.							
Firstborn (=1)	0.032	0.032	0.040	0.047	0.040	0.033	0.038
	(0.006)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
Mean of Y	0.039	0.048	0.021	0.023	0.016	–0.017	0.012
# of Households	144,991	269,384	473,647	717,637	868,932	1,084,510	3,559,101
Observations	289,982	538,768	947,294	1,435,274	1,737,864	2,169,020	7,118,202
D. Family fixed effects. Within-family controls.							
Firstborn (=1)	0.032	0.035	0.040	0.050	0.042	0.038	0.041
	(0.006)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.001)
Mean of Y	0.039	0.048	0.021	0.023	0.016	–0.017	0.012
# of Households	144,989	269,380	473,647	717,637	868,932	1,084,510	3,559,095
Observations	289,978	538,760	947,294	1,435,274	1,737,864	2,169,020	7,118,190

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include year of birth fixed effects. Panel B controls for race and adds fixed effects at the household level for family size and father's occupational status in addition to individual-level fixed effects for state of birth and mother's age at birth. We use OCC1950 codes to assign the father's occupation to one of four categories: white collar, skilled, farmers, and unskilled. See Section 3 for details. Panel D includes state of birth and mother's age at birth FEs. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A2. Robustness of Firstborn Premium to 40-year gap

	(1) 1860 to 1900	(2) 1870 to 1910	(3) 1880 to 1920	(4) 1900 to 1940	(5) Pooled
A. OLS. No controls except birth year fixed effects.					
Firstborn (=1)	0.063	0.037	0.055	0.040	0.047
	(0.004)	(0.003)	(0.003)	(0.002)	(0.001)
Mean of Y	0.132	0.184	0.127	0.211	0.173
# of Households	195,874	275,876	403,650	646,010	1,521,410
Observations	391,748	551,752	807,300	1,292,020	3,042,820
B. OLS. Family and individual controls.					
Firstborn (=1)	0.036	0.027	0.039	0.022	0.029
	(0.004)	(0.004)	(0.003)	(0.002)	(0.001)
Mean of Y	0.132	0.184	0.127	0.211	0.173
# of Households	195,874	275,876	403,650	646,010	1,521,410
Observations	391,747	551,751	807,299	1,292,020	3,042,817
C. Family fixed effects. No controls except birth year fixed effects.					
Firstborn (=1)	0.037	0.024	0.043	0.030	0.033
	(0.006)	(0.005)	(0.004)	(0.003)	(0.002)
Mean of Y	0.132	0.184	0.127	0.211	0.173
# of Households	195,874	275,876	403,650	646,010	1,521,410
Observations	391,748	551,752	807,300	1,292,020	3,042,820
D. Family fixed effects. Within-family controls.					
Firstborn (=1)	0.039	0.024	0.043	0.031	0.034
	(0.006)	(0.005)	(0.004)	(0.003)	(0.002)
Mean of Y	0.132	0.184	0.127	0.211	0.173
# of Households	195,873	275,875	403,649	646,010	1,521,407
Observations	391,746	551,750	807,298	1,292,020	3,042,814

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include year of birth fixed effects. Panel B controls for race and adds fixed effects at the household level for family size and father's occupational status in addition to individual-level fixed effects for state of birth and mother's age at birth. We use OCC1950 codes to assign the father's occupation to one of four categories: white collar, skilled, farmers, and unskilled. See Section 3 for details. Panel D includes state of birth and mother's age at birth FEs. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A3. Robustness of Firstborn Premium

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. Restrict age range of kids in baseline census, 0 to 12.						
Firstborn (=1)	0.033	0.032	0.042	0.037	0.029	0.034
	(0.008)	(0.006)	(0.004)	(0.003)	(0.003)	(0.002)
Mean of Y	0.190	0.178	0.199	0.209	0.179	0.192
# of Households	121,944	208,800	356,213	539,667	650,878	1,877,502
Observations	243,888	417,600	712,426	1,079,334	1,301,756	3,755,004
B. Restrict to kids age 0 to 12 and mothers age 18 to 30.						
Firstborn (=1)	0.040	0.037	0.038	0.041	0.031	0.037
	(0.010)	(0.007)	(0.005)	(0.004)	(0.003)	(0.002)
Mean of Y	0.197	0.193	0.222	0.216	0.181	0.201
# of Households	85,599	143,906	249,043	377,955	463,550	1,320,053
Observations	171,198	287,812	498,086	755,910	927,100	2,640,106
C. Firstborn vs. all laterborns.						
Firstborn (=1)	0.028	0.031	0.042	0.042	0.031	0.036
	(0.006)	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)
Mean of Y	0.170	0.160	0.178	0.183	0.162	0.171
# of Households	164,142	295,258	474,942	722,628	874,222	2,531,192
Observations	393,725	698,924	1,137,916	1,728,708	2,079,280	6,038,553
D. Drop families with exactly two children.						
Firstborn (=1)	0.033	0.034	0.038	0.040	0.026	0.034
	(0.008)	(0.005)	(0.004)	(0.003)	(0.003)	(0.002)
Mean of Y	0.172	0.159	0.171	0.166	0.140	0.158
# of Households	131,871	229,916	373,306	540,585	625,259	1,900,937
Observations	263,742	459,832	746,612	1,081,170	1,250,518	3,801,874
E. Drop families if younger sibling is youngest in household.						
Firstborn (=1)	0.035	0.028	0.036	0.038	0.021	0.031
	(0.009)	(0.006)	(0.005)	(0.004)	(0.003)	(0.002)
Mean of Y	0.168	0.149	0.164	0.157	0.138	0.152
# of Households	101,054	176,319	287,559	415,656	480,599	1,461,187
Observations	202,108	352,638	575,118	831,312	961,198	2,922,374

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample, with the exception of Panel C where the comparison is to all siblings who can be included in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects – corresponding to the specification in Panel D of Table 1. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A4. Firstborn Premium: Firstborn Male with Older Sister

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. Family fixed effects. Firstborn boy is 2nd born.						
Firstborn (=1)	–0.010	–0.001	0.013	0.010	0.011	0.008
	(0.010)	(0.007)	(0.006)	(0.004)	(0.004)	(0.002)
Mean of Y	0.130	0.136	0.128	0.128	0.093	0.118
# of Households	74,294	119,400	188,495	263,331	300,367	945,887
Observations	148,588	238,800	376,990	526,662	600,734	1,891,774

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the second-born (male) sibling with an older sister compared to the younger (male) sibling closest in birth order whom we can include in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects – corresponding to the specification in Panel D of Table 1. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A5. Firstborn Premium: Above vs. Below Median Family Size

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. Family fixed effects. Within-family controls. Small family size (up to 3 children)						
Firstborn (=1)	0.019	0.027	0.031	0.037	0.026	0.030
	(0.011)	(0.007)	(0.006)	(0.004)	(0.003)	(0.002)
Mean of Y	0.249	0.218	0.267	0.277	0.254	0.259
# of Households	71,765	140,050	226,163	380,517	492,930	1,311,425
Observations	143,530	280,100	452,326	761,034	985,860	2,622,850
B. Family fixed effects. Within-family controls. Large family size (3-15 children)						
Firstborn (=1)	0.026	0.035	0.042	0.041	0.027	0.035
	(0.009)	(0.007)	(0.005)	(0.004)	(0.004)	(0.002)
Mean of Y	0.143	0.133	0.132	0.119	0.092	0.117
# of Households	92,373	155,208	248,779	342,110	381,292	1,219,762
Observations	184,746	310,416	497,558	684,220	762,584	2,439,524

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. Panel A restricts the sample to sibships with 2 to 3 siblings (below median family size) enumerated in the census during childhood, while Panel B restricts to sibships with 4 to 15 siblings (above median family size). The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects – corresponding to the specification in Panel D of Table 1. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A6. Firstborn Premium: Above vs. Below Mother's Age at First Birth

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. Family fixed effects. Within-family controls. Below median mother's age at first birth.						
Firstborn (=1)	0.037	0.028	0.036	0.043	0.022	0.032
	(0.010)	(0.007)	(0.006)	(0.004)	(0.004)	(0.002)
Mean of Y	0.149	0.132	0.145	0.139	0.107	0.130
# of Households	94,330	162,708	260,511	358,915	437,384	1,313,848
Observations	188,660	325,416	521,022	717,830	874,768	2,627,696
B. Family fixed effects. Within-family controls. Above median mother's age at first birth.						
Firstborn (=1)	0.013	0.029	0.036	0.033	0.025	0.029
	(0.011)	(0.008)	(0.006)	(0.004)	(0.004)	(0.002)
Mean of Y	0.246	0.229	0.262	0.274	0.267	0.263
# of Households	66,980	127,584	206,983	340,934	412,528	1,155,009
Observations	133,960	255,168	413,966	681,868	825,056	2,310,018

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. The table restricts the sample to households with below (Panel A) and above (Panel B) median mother's age at first birth. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects – corresponding to the specification in Panel D of Table 1. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A7. Firstborn Premium for Occupational Education Percentile

		Split by race		Split by father's occupation			
	(1) Full Sample	(2) White	(3) Black	(4) Unskilled	(5) Farmer	(6) Skilled	(7) White Collar
A. OLS. No controls except birth year fixed effects.							
Firstborn (=1)	0.863	0.954	0.241	0.483	0.507	1.299	1.074
	(0.024)	(0.024)	(0.097)	(0.051)	(0.035)	(0.065)	(0.064)
Mean of Y	54.838	56.722	15.223	52.030	49.556	61.923	71.470
# of Households	2,528,936	2,414,084	114,852	565,425	1,064,455	317,024	302,882
Observations	5,057,872	4,828,168	229,704	1,130,850	2,128,910	634,048	605,764
B. OLS. Family and individual controls.							
Firstborn (=1)	0.942	0.971	0.166	0.862	0.695	1.133	1.393
	(0.024)	(0.025)	(0.104)	(0.054)	(0.037)	(0.070)	(0.069)
Mean of Y	54.838	56.722	15.222	52.030	49.556	61.923	71.470
# of Households	2,528,936	2,414,084	114,852	565,425	1,064,455	317,024	302,881
Observations	5,057,872	4,828,168	229,695	1,130,844	2,128,908	634,039	605,751
C. Family fixed effects. No controls except birth year fixed effects.							
Firstborn (=1)	1.021	1.049	0.105	0.975	0.973	1.377	1.029
	(0.033)	(0.034)	(0.140)	(0.072)	(0.049)	(0.095)	(0.095)
Mean of Y	54.838	56.722	15.223	52.030	49.556	61.923	71.470
# of Households	2,528,936	2,414,084	114,852	565,425	1,064,455	317,024	302,882
Observations	5,057,872	4,828,168	229,704	1,130,850	2,128,910	634,048	605,764
D. Family fixed effects. Within-family controls.							
Firstborn (=1)	1.014	1.032	0.085	0.990	1.001	1.309	0.974
	(0.034)	(0.035)	(0.145)	(0.075)	(0.050)	(0.098)	(0.098)
Mean of Y	54.838	56.722	15.221	52.030	49.556	61.923	71.470
# of Households	2,528,936	2,414,084	114,843	565,418	1,064,453	317,014	302,870
Observations	5,057,872	4,828,168	229,686	1,130,836	2,128,906	634,028	605,740

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the "adjusted Song score" which can be interpreted as an occupational education percentile score. We follow the implementation in Ward (2023) who modifies a similar measure in Song et al. (2019). The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All columns are estimated on a sample that pools observations across all linked census waves with 30-year gaps. Specifications in Panels A to D correspond to specifications from column 6 of Table 1. Column 1 includes the full sample. Columns 2 to 3 include only the sub-samples of White and Black families, respectively. Columns 4 to 7 split the sample by father's occupational status in the baseline census: unskilled, farmer, skilled, and white collar, respectively. Standard errors are clustered at the family level.

Table A8. Firstborn Premium for Occupational and Social Outcomes of Black Men

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. Std. OCCSCORE						
Firstborn (=1)	–0.017	0.008	0.010	0.001	–0.001	0.003
	(0.047)	(0.015)	(0.011)	(0.010)	(0.009)	(0.005)
Mean of Y	–0.089	–0.326	–0.379	–0.406	–0.450	–0.400
# of Households	1,680	14,647	27,632	30,558	40,612	115,129
Observations	3,360	29,294	55,264	61,116	81,224	230,258
B. P(Single)						
Firstborn (=1)	0.028	–0.016	–0.000	–0.008	–0.008	–0.007
	(0.022)	(0.007)	(0.005)	(0.004)	(0.004)	(0.002)
Mean of Y	0.168	0.141	0.115	0.111	0.147	0.129
# of Households	1,770	17,073	29,849	33,158	45,237	127,087
Observations	3,540	34,146	59,698	66,316	90,474	254,174
C. P(Any children)						
Firstborn (=1)	–0.018	0.011	–0.005	0.007	0.004	0.003
	(0.030)	(0.009)	(0.007)	(0.007)	(0.006)	(0.003)
Mean of Y	0.656	0.660	0.648	0.565	0.509	0.579
# of Households	1,770	17,073	29,835	33,158	45,237	127,073
Observations	3,540	34,146	59,670	66,316	90,474	254,146
D. Number of children — any kids						
Firstborn (=1)	–0.139	0.047	0.067	0.032	–0.002	0.031
	(0.185)	(0.064)	(0.050)	(0.053)	(0.049)	(0.026)
Mean of Y	3.541	3.864	3.755	3.560	3.375	3.613
# of Households	848	8,094	13,981	12,122	13,882	48,927
Observations	1,696	16,188	27,962	24,244	27,764	97,854

Notes: Each coefficient comes from a separate regression on Black male-only samples. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects – corresponding to the specification in Panel D of Table 1. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A9. Firstborn Premium for Occupational and Social Outcomes of White Men

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. Std. OCCSCORE						
Firstborn (=1)	0.025	0.031	0.038	0.039	0.027	0.033
	(0.007)	(0.005)	(0.004)	(0.003)	(0.002)	(0.002)
Mean of Y	0.192	0.199	0.232	0.229	0.214	0.218
# of Households	162,460	280,610	447,309	692,067	833,609	2,416,055
Observations	324,920	561,220	894,618	1,384,134	1,667,218	4,832,110
B. P(Single)						
Firstborn (=1)	–0.019	–0.019	–0.022	–0.022	–0.021	–0.021
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Mean of Y	0.132	0.185	0.170	0.140	0.148	0.153
# of Households	171,305	336,952	489,372	760,471	886,627	2,644,727
Observations	342,610	673,904	978,744	1,520,942	1,773,254	5,289,454
C. P(Any children)						
Firstborn (=1)	0.019	0.016	0.020	0.021	0.021	0.020
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Mean of Y	0.759	0.676	0.674	0.677	0.645	0.671
# of Households	171,305	336,935	489,036	760,469	886,627	2,644,372
Observations	342,610	673,870	978,072	1,520,938	1,773,254	5,288,744
D. Number of children — any kids						
Firstborn (=1)	0.117	0.090	0.076	0.078	0.057	0.076
	(0.014)	(0.011)	(0.009)	(0.007)	(0.006)	(0.004)
Mean of Y	3.462	3.234	3.114	2.797	2.547	2.888
# of Households	107,694	175,708	254,921	394,861	422,699	1,355,883
Observations	215,388	351,416	509,842	789,722	845,398	2,711,766

Notes: Each coefficient comes from a separate regression on White male-only samples. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects – corresponding to the specification in Panel D of Table 1. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A10. Firstborn Premium: Comparisons with Specific Sibling Order

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. Family fixed effects. Within-family controls. 1st vs. 2nd.						
Firstborn (=1)	0.019	0.031	0.040	0.041	0.026	0.033
	(0.009)	(0.006)	(0.005)	(0.003)	(0.003)	(0.002)
Mean of Y	0.226	0.187	0.225	0.234	0.222	0.223
# of Households	91,954	175,005	284,816	456,052	574,243	1,582,070
Observations	183,908	350,010	569,632	912,104	1,148,486	3,164,140
B. Family fixed effects. Within-family controls. 1st vs. 3rd.						
Firstborn (=1)	0.048	0.065	0.056	0.048	0.053	0.053
	(0.016)	(0.011)	(0.009)	(0.006)	(0.006)	(0.004)
Mean of Y	0.175	0.160	0.177	0.176	0.149	0.165
# of Households	61,064	108,132	180,658	271,491	319,650	940,995
Observations	122,128	216,264	361,316	542,982	639,300	1,881,990
C. Family fixed effects. Within-family controls. 1st vs. 4th.						
Firstborn (=1)	0.043	0.064	0.071	0.055	0.053	0.058
	(0.026)	(0.019)	(0.015)	(0.011)	(0.010)	(0.006)
Mean of Y	0.131	0.123	0.127	0.118	0.084	0.111
# of Households	37,813	63,626	105,739	150,711	169,983	527,872
Observations	75,626	127,252	211,478	301,422	339,966	1,055,744

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the second-born (Panel A), third-born (Panel B) or fourth-born (Panel C) male sibling. All regressions include family, year of birth, state of birth, and mother's age at birth fixed effects. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A11. Heterogeneity by Father's Occupational Status

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. Father's occupation is white collar.						
Firstborn (=1)	0.017	0.007	0.029	0.024	0.021	0.022
	(0.015)	(0.009)	(0.007)	(0.005)	(0.004)	(0.003)
Mean of Y	0.324	0.217	0.257	0.276	0.199	0.240
# of Households	24,878	65,208	107,765	163,365	201,850	563,066
Observations	49,756	130,416	215,530	326,730	403,700	1,126,132
B. Father's occupation is skilled.						
Firstborn (=1)	0.020	0.035	0.046	0.042	0.023	0.034
	(0.009)	(0.007)	(0.005)	(0.004)	(0.004)	(0.002)
Mean of Y	−0.026	−0.072	−0.074	−0.161	−0.171	−0.122
# of Households	91,598	135,943	235,351	291,445	310,880	1,065,217
Observations	183,196	271,886	470,702	582,890	621,760	2,130,434
C. Father's occupation is farmer.						
Firstborn (=1)	0.037	0.024	0.025	0.018	0.029	0.025
	(0.017)	(0.014)	(0.011)	(0.007)	(0.006)	(0.004)
Mean of Y	0.461	0.514	0.570	0.549	0.448	0.505
# of Households	23,894	34,332	53,373	88,341	116,562	316,502
Observations	47,788	68,664	106,746	176,682	233,124	633,004
D. Father's occupation is unskilled.						
Firstborn (=1)	0.068	0.058	0.038	0.058	0.031	0.044
	(0.032)	(0.022)	(0.015)	(0.009)	(0.007)	(0.005)
Mean of Y	0.934	0.898	0.961	0.808	0.724	0.809
# of Households	12,960	23,204	49,179	84,614	130,937	300,894
Observations	25,920	46,408	98,358	169,228	261,874	601,788

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. Each panel presents estimates using the subsample of individuals whose fathers were employed in the stated occupational category in the baseline census. We use OCC1950 codes to assign the father's occupation to one of four categories: white collar (Panel A), skilled (Panel B), farmers (Panel C), and unskilled (Panel D). White-collar occupations include professionals (OCC1950 codes 0-99), managers (200-299), clerical workers (300-399), and sales workers (400-499). Skilled workers are craftsmen (500-599). Farmers include farm owners and tenants (OCC1950 code 100) in addition to farm managers (OCC1950 code 123). Unskilled workers include operatives (600-699), service workers (700-799), farm laborers and general laborers (800-970). The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects – corresponding to the specification in Panel D of Table 1. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A12. Robustness of Firstborn Premium to Linking Method

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. Baseline. Family fixed effects. Within-family controls.						
Firstborn (=1)	0.025	0.030	0.037	0.038	0.026	0.032
	(0.007)	(0.005)	(0.004)	(0.003)	(0.002)	(0.001)
Mean of Y	0.189	0.173	0.196	0.202	0.183	0.190
# of Households	164,142	295,258	474,942	722,628	874,222	2,531,192
Observations	328,284	590,516	949,884	1,445,256	1,748,444	5,062,384
B. Links included in Census Linking Project.						
Firstborn (=1)	0.011	0.042	0.046	0.044	0.032	0.037
	(0.016)	(0.012)	(0.009)	(0.006)	(0.005)	(0.003)
Mean of Y	0.210	0.208	0.259	0.246	0.205	0.226
# of Households	26,462	42,254	77,335	155,689	217,223	518,963
Observations	52,924	84,508	154,670	311,378	434,446	1,037,926
C. Links generated by XGBoost ML algorithm.						
Firstborn (=1)	0.028	0.044	0.044	0.040	0.030	0.036
	(0.012)	(0.008)	(0.006)	(0.004)	(0.003)	(0.002)
Mean of Y	0.212	0.186	0.214	0.232	0.196	0.209
# of Households	51,654	88,806	177,444	272,285	388,472	978,661
Observations	103,308	177,612	354,888	544,570	776,944	1,957,322
D. Links on Family Tree (i.e. identified by FamilySearch users).						
Firstborn (=1)	0.043	0.038	0.038	0.046	0.025	0.036
	(0.011)	(0.008)	(0.006)	(0.004)	(0.004)	(0.002)
Mean of Y	0.088	0.069	0.135	0.133	0.136	0.125
# of Households	54,597	85,956	203,520	251,226	307,218	902,517
Observations	109,194	171,912	407,040	502,452	614,436	1,805,034

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include family, year of birth, mother's age at birth, and state of birth fixed effects. Panel A repeats the estimates in Panel D of Table 1. The subsequent panels restrict the sample to the subset of families for which the links of the firstborn child (a boy) and at least one later born brother are included in the Census Linking Project (panel B), are generated by the XGBoost ML algorithm (panel C), or appear on the Family Tree, i.e. are identified by FamilySearch users (panel D). Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Standard errors are clustered at the family level.

Table A13. Firstborn Premium for Occupation Status of Men, Weighted Estimates

	(1) 1850 to 1880	(2) 1870 to 1900	(3) 1880 to 1910	(4) 1900 to 1930	(5) 1910 to 1940	(6) Pooled
A. OLS. No controls except birth year fixed effects.						
Firstborn (=1)	0.034	0.012	0.034	0.025	0.018	0.023
	(0.005)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)
Mean of Y	0.218	0.152	0.148	0.162	0.120	0.151
# of Households	164,142	295,257	474,940	722,628	874,222	2,531,189
Observations	328,284	590,514	949,880	1,445,256	1,748,444	5,062,378
B. OLS. Family and individual controls.						
Firstborn (=1)	0.020	0.019	0.029	0.024	0.017	0.022
	(0.005)	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)
Mean of Y	0.218	0.152	0.148	0.162	0.120	0.151
# of Households	164,142	295,257	474,940	722,628	874,222	2,531,189
Observations	328,284	590,514	949,880	1,445,256	1,748,444	5,062,378
C. Family fixed effects. No controls except birth year fixed effects.						
Firstborn (=1)	0.023	0.023	0.030	0.032	0.019	0.026
	(0.007)	(0.005)	(0.004)	(0.003)	(0.003)	(0.002)
Mean of Y	0.218	0.152	0.148	0.162	0.120	0.151
# of Households	164,142	295,257	474,940	722,628	874,222	2,531,189
Observations	328,284	590,514	949,880	1,445,256	1,748,444	5,062,378
D. Family fixed effects. Within-family controls.						
Firstborn (=1)	0.022	0.026	0.032	0.034	0.023	0.028
	(0.007)	(0.005)	(0.004)	(0.003)	(0.003)	(0.002)
Mean of Y	0.218	0.152	0.148	0.162	0.120	0.151
# of Households	164,142	295,257	474,940	722,628	874,222	2,531,189
Observations	328,284	590,514	949,880	1,445,256	1,748,444	5,062,378

Notes: Each coefficient comes from a separate regression on male-only samples. The dependent variable is the OCCSCORE (Ruggles et al. 2023) of an individual's occupation standardized separately for the male population in each outcome year census. The coefficient on the indicator variable firstborn captures the effect on the outcome of being the firstborn sibling compared to the sibling closest in birth order whom we can include in the sample. All regressions include year of birth fixed effects. Panel B controls for race and adds fixed effects at the household level for family size and father's occupational status in addition to individual-level fixed effects for state of birth and mother's age at birth. We use OCC1950 codes to assign the father's occupation to one of four categories: white collar, skilled, farmers, and unskilled. See Section 3 for details. Panels C and D include family FEs. Panel D additionally includes state of birth and mother's age at birth FEs. Columns 1 to 5 present estimates based on separate linked census pairs. Column 6 pools together all census waves and interacts all control variables with census wave FEs. Households are weighted to adjust the sample for representativeness with respect to the baseline population of interest in the childhood census (Bailey et al. 2020). Household weights are given by $w = q \cdot \hat{p}/(1-\hat{p})$ where q is the share of observations in the baseline population that are included in the estimation sample and $\hat{p}$ is the predicted probability of being part of the estimation sample as a function of race and interactions of race with 10-year bins of own age and mother's age at first birth as well as fixed effects for census region, father's occupational status (See Table A11), and family size. Standard errors are clustered at the family level.