# INTERGENERATIONAL CORRELATIONS IN LONGEVITY

Sandra E. Black (Columbia University)
Neil Duzett (Texas A&M)
Adriana Lleras-Muney (UCLA)
Nolan Pope (University of Maryland)
Joseph Price (BYU)

July 28, 2022

# Motivation

- Substantial literature in economics on intergenerational correlations in socio-economic status
  - Income, education, and wealth (Solon 1992; Black and Devereux 2011; Stuhler 2018)
- Burgeoning literature on intergenerational correlations of health
  - Mother's and child's birth weight (Currie and Moretti 2007; Giuntella et al. 2022)
  - Self-reported health status, health indices, or specific conditions (Halliday et al. 2019, 2021; Andersen 2019; Thompson 2014)
- Longevity is a large component of welfare (Jones and Klenow 2016)
  - How correlated is longevity across generations?

# What We Do

- Estimate the intergenerational correlation in longevity (IGCL) using new data:
  - Matches family trees available from Family Search to Census data
  - Large sample: over 8 million individuals born between 1880-1920
- Compare IGCL to:
  - IGCL of grandparents
  - Sibling and twin correlations in longevity
  - Correlations in socio-economic status
- Investigate:
  - The role of environment, genetics, and SES

# Contribution to Literature

**Economics:**

- Little research examining intergenerational correlations in longevity
  - Longevity is a good summary measure of lifetime health
  - Strongly correlated with education, occupation & other SES measures, but contains additional information

- Longevity has advantages over other measures (SES, SRHS, conditions)
  - Observed for men and women
  - Easily computed
  - More comparable across time and location
  - Used across multiple generation

**Demography:**

- Establish representativeness of the sample
- Substantially larger sample: can investigate subgroup heterogeneity
- Compare correlations in longevity to correlations in other economic outcomes

# Data

- Census data from 1900, 1910, and 1920
  - Focus on everyone who was born between the years 1880 and 1920
  - Also link to 1940 Census

- Family Search Family Tree data
  - Largest collection in world (1.3 billion individuals represented)
  - Profiles provide information on date of birth, date of death, and therefore longevity of children, parents, and grandparents
  - Match to Census using records attached to profiles on Family Tree

- Main sample: Over 8 million individuals
  - Both child and parent have data and lived past age 25

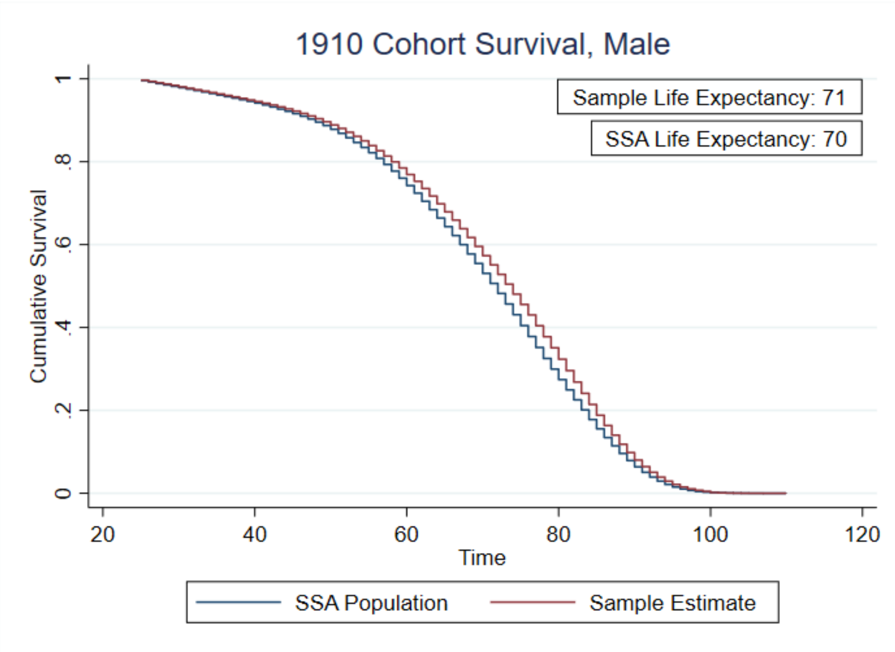
- Data issues
  - Who is represented?
  - Is the age at death data accurate?

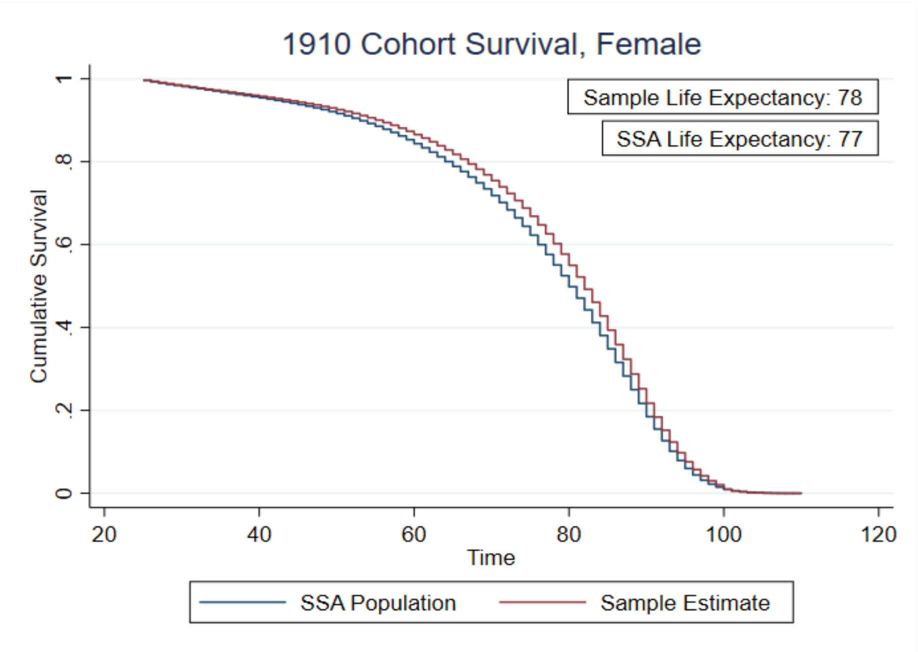
# Representativeness of Data

| Variable | Matched Analysis Data | Matched Analysis Data (Weighted) | Full Census Sample |
|---|---|---|---|
| Female | 0.458 | 0.460 | 0.444 |
| Birth Year | 1900 | 1893 | 1892 |
| White | 0.994 | 0.993 | 0.863 |
| Black | 0.004 | 0.005 | 0.128 |
| Northeast | 0.119 | 0.304 | 0.304 |
| Midwest | 0.411 | 0.299 | 0.299 |
| South | 0.351 | 0.301 | 0.301 |
| West | 0.118 | 0.082 | 0.082 |
| Immigrant | 0.011 | 0.029 | 0.190 |
| Father is Immigrant | 0.138 | 0.161 | 0.378 |
| Mother is Immigrant | 0.115 | 0.139 | 0.357 |
| Observations | 8,623,936 | 8,623,936 | 38,947,264 |

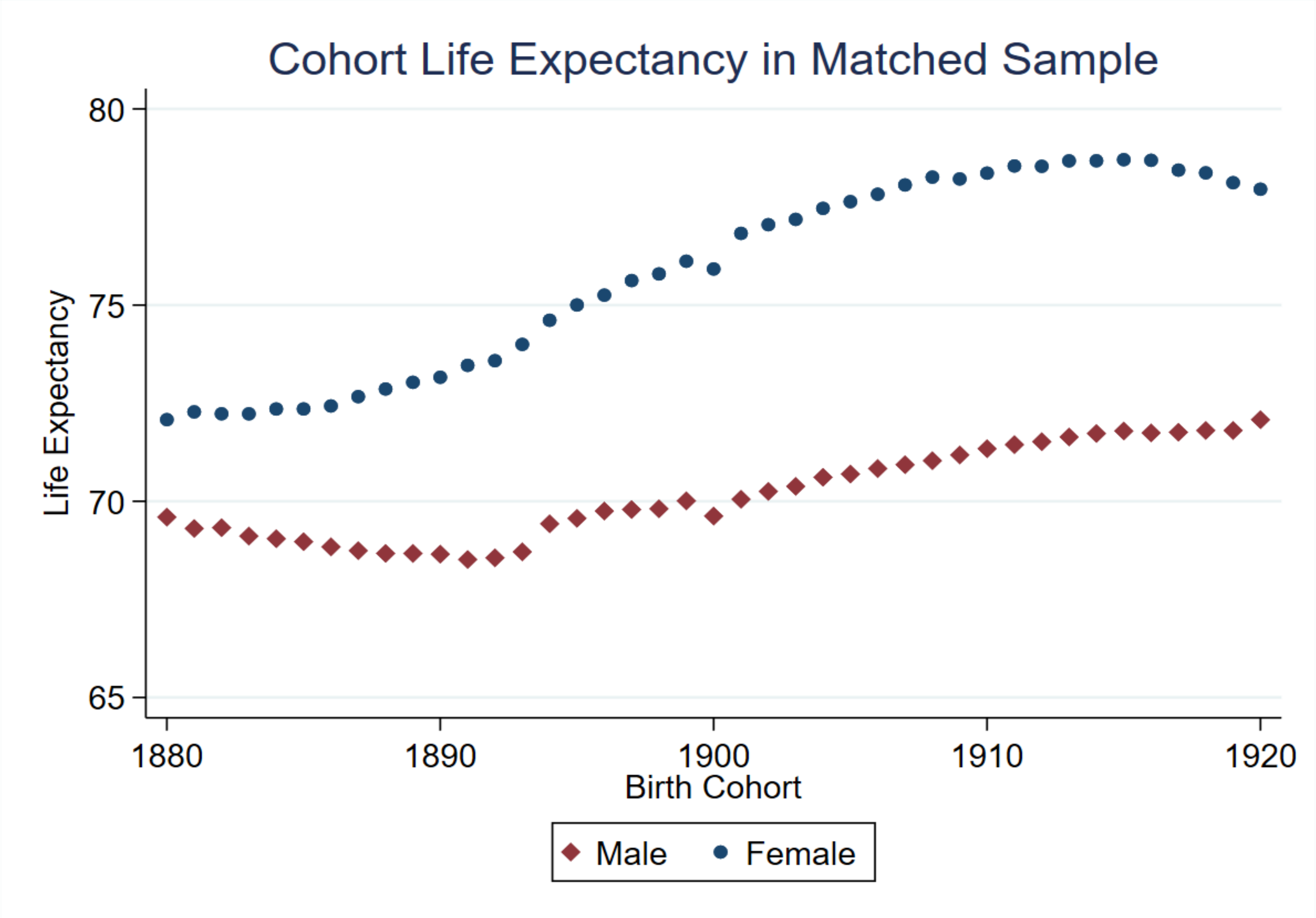
- Females well represented: improvement relative to other linked records
- Blacks and immigrants vastly under-represented: similar to other genealogical research
- Geographic distribution skewed towards Midwest; NE under-represented
- Older cohorts under-represented

# Survival Profiles




- Individuals in our data live longer. Our data likely biased towards long lived, but SSA data also not clearly representative. Histograms reveals little [age-heaping]
- Similar for other cohorts




- Gender gap and its evolution consistent with prior research
- Decline in LE for males in 19th century maybe due to urbanization, effects of civil war (similar to trends in men's height)

# Data Generating Model

$$L_{ijsc}^{g} = \delta\,\frac{1}{2}\left[G_{jc-1}^{f} + \eta_{ijc}^{f} + G_{jsc-1}^{m} + \eta_{ijc}^{m}\right] + \alpha_{sc} + \gamma_{i}^{g} + \theta_{j}^{SES} + \varepsilon_{i}$$

- Longevity of individual *i* born in family *j* in state *s* and cohort/year *c* depends on
  - Genes: individuals get a random ½ from each parent ($G_{jsc-1}^{f}$, $G_{jsc-1}^{m}$) with some genetic variation ($\eta_{ijc}^{f}, \eta_{ijc}^{m}$)
  - Environmental factors: technology, water/air quality, etc. ($\alpha_{sc}$)
  - SES: education, income, wealth, rank, etc. ($\theta_{j}^{SES}$)
  - Gender: genes, hormones, social effects ($\gamma_{i}^{g}$)

- Strong assumptions
  - Linearity. No gene-environment interactions. No behavior/optimization
  - Parental lifespan itself has no effect on child lifespan

# Model Predictions

- IGCL captures:
  - Genetic component, including assortative mating
  - Environmental component
  - SES component

- If $\alpha_{sc} = \alpha_s$ (environment is stable) then:
  - $\beta_{twins} > \beta_{brothers} = \beta_{father-son} > \beta_{mother-son} > \beta_{father-mother}$
  - $\beta_{parent\ average} > \beta_{one\ parent}$
  - $\beta_{father} > \beta_{grandfather}$

# Estimation Strategy

$$L_i^C = \beta_0 + \beta_1 L_i^P + X\beta_2 + e_{ijc}$$

- $L_i^C$: age at death of individual *i,* conditional on surviving to age 25
- $L_i^P$: age at death of parent, conditional on surviving to age 25
- X is a set of control variables
  - Basic model controls for birth cohort for parents and children to account for longevity trends

- Robustness checks:
  - Different functional form
  - Additional controls
  - Linearity assumption
  - Cutoff choice (25, 35, etc.)

# Results with Varying Controls

| Model | No Controls | + Parent and Child Birth Year FE | + Parent and Child State of Birth FE | + Race and birth order FE |
|---|---|---|---|---|
| Son/Father | 0.104 | 0.109 | 0.106 | 0.106 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Daughter/Mother | 0.088 | 0.088 | 0.086 | 0.086 |
| | (0.001) | (0.001) | (0.001) | (0.001) |

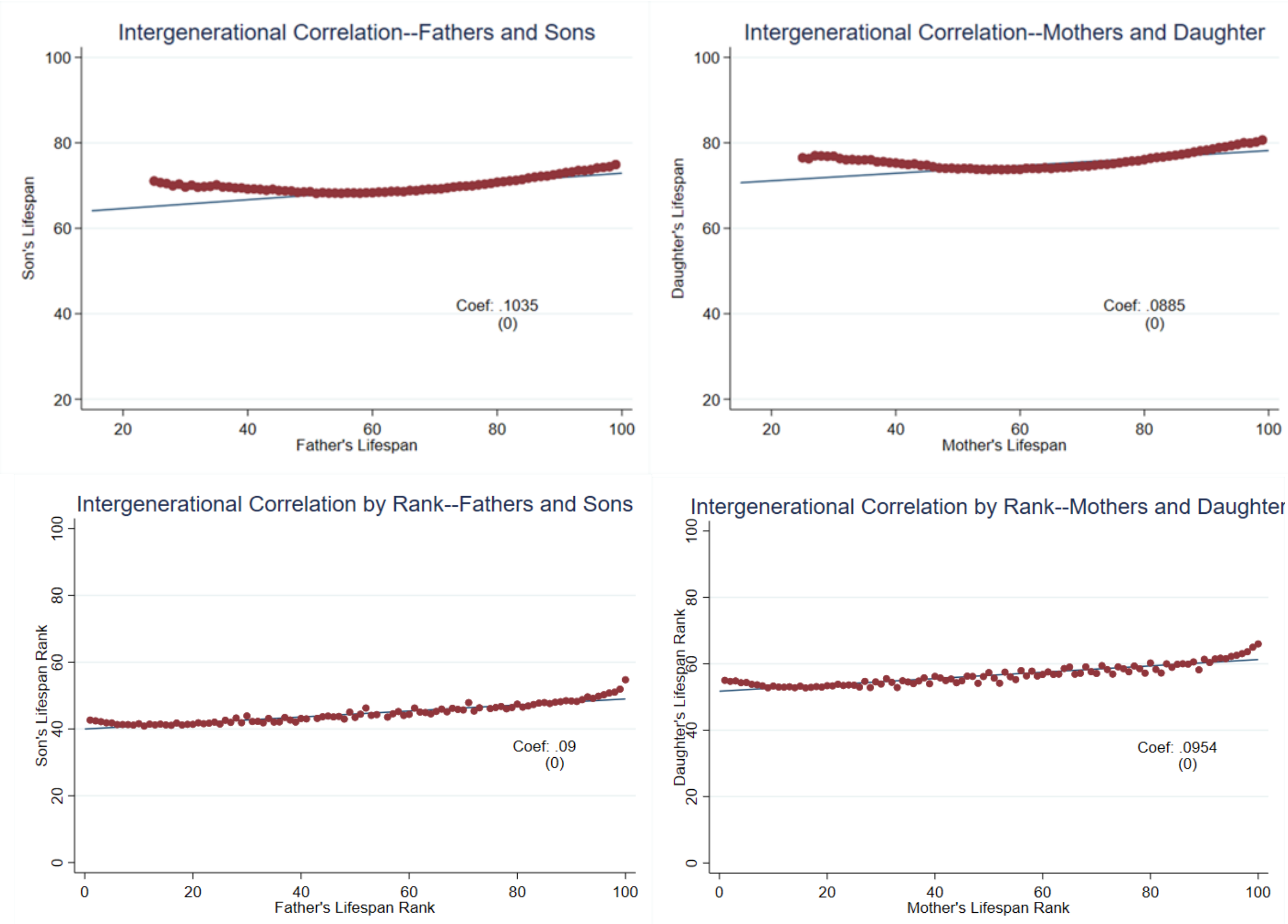
- Not sensitive to controls for birth year, birth state, race, or birth order
- IGCL tends to be larger for males than females

# Results by Gender and Outcome

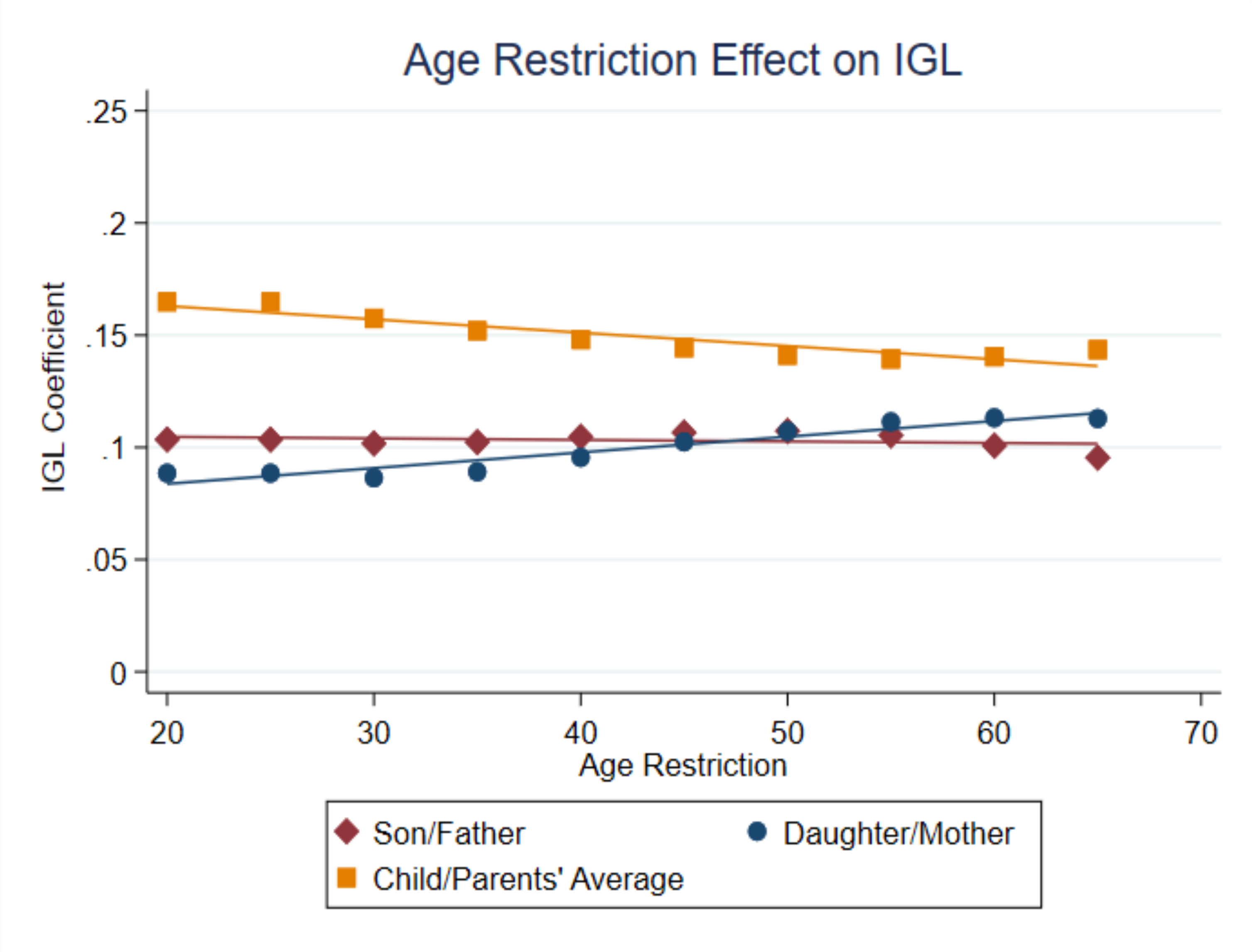
| Model | Outcome | | | |
|---|---|---|---|---|
| | Lifespan (Years) | Lifespan (Weighted*) | Percentile | Log Lifespan |
| Son/Father | 0.109 | 0.121 | 0.092 | 0.101 |
| | (0.001) | (0.001) | (0.000) | (0.001) |
| Son/Mother | 0.075 | 0.087 | 0.080 | 0.064 |
| | (0.001) | (0.001) | (0.000) | (0.001) |
| Son/Parents' Average | 0.170 | 0.192 | 0.164 | 0.168 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Daughter/Father | 0.090 | 0.104 | 0.079 | 0.079 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Daughter/Mother | 0.088 | 0.106 | 0.095 | 0.074 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Daughter/Parents' Average | 0.170 | 0.199 | 0.167 | 0.163 |
| | (0.001) | (0.002) | (0.001) | (0.001) |
| Father/Mother | 0.041 | 0.044 | 0.046 | 0.039 |
| | (0.001) | (0.001) | (0.001) | (0.001) |

- Reweighting and function form has limit impact on IGCL
- Verify $\beta_{father-son} > \beta_{mother-son} > \beta_{father-mother}$, and $\beta_{parent\ average} > \beta_{one\ parent}$
- Assortative mating based on longevity low

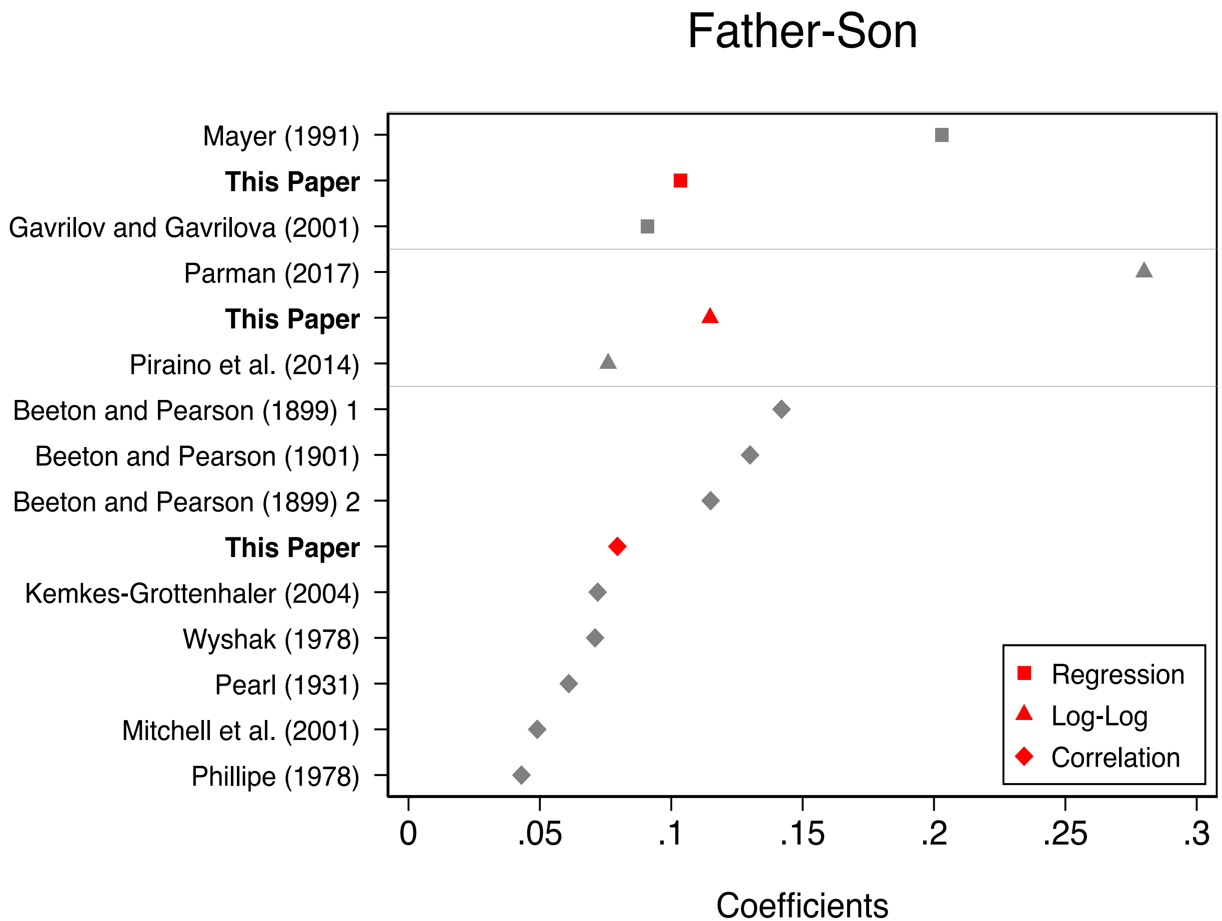
# Testing for Linearity



# Sensitivity to Truncation Age



# Comparison to Previous Estimates




- Despite differences in sample, time period, and truncation, estimates are comparable, and all in the low range.

# Multiple Generations

| | Bivariate Regressions | | Multiple Regressions | |
|---|---|---|---|---|
| Variable | Son's Lifespan | Daughter's Lifespan | Son's Lifespan | Daughter's Lifespan |
| Father's Lifespan | 0.090 | 0.062 | 0.083 | 0.055 |
| | (0.002) | (0.003) | (0.002) | (0.003) |
| Mother's Lifespan | 0.068 | 0.076 | 0.062 | 0.071 |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Paternal Grandfather's Lifespan | 0.031 | 0.023 | 0.025 | 0.017 |
| | (0.002) | (0.003) | (0.002) | (0.003) |
| Paternal Grandmother's Lifespan | 0.020 | 0.022 | 0.014 | 0.018 |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Maternal Grandfather's Lifespan | 0.033 | 0.027 | 0.028 | 0.021 |
| | (0.002) | (0.003) | (0.002) | (0.003) |
| Maternal Grandmother's Lifespan | 0.015 | 0.019 | 0.010 | 0.014 |
| | (0.002) | (0.002) | (0.002) | (0.002) |
| Observations | 253,035 | 206,980 | 253,035 | 206,980 |

- $\beta_{father} > \beta_{grandfather}$ as model predicts
- Grandparents still matter controlling for parents

# Sibling Correlations

| | Siblings | | | |
|---|---|---|---|---|
| Outcome: | Lifespan | Education | HH Income | IGCLF |
| Brother/Brother | 0.139 | 0.537 | 0.248 | 0.095 |
| | (0.001) | (0.001) | (0.009) | (0.001) |
| N | 885,041 | 885,041 | 885,041 | 1,221,808 |
| Sister/Sister | 0.105 | 0.577 | 0.256 | 0.079 |
| | (0.002) | (0.002) | (0.012) | (0.001) |
| N | 480,144 | 480,144 | 480,144 | 807,598 |
| Sister/Brother | 0.040 | 0.508 | 0.209 | 0.087 |
| | (0.001) | (0.001) | (0.008) | (0.001) |
| N | 1,242,108 | 1.242.108 | 1,242,108 | 1,925,854 |

- Sibling correlations in lifespan somewhat larger than the IGCL with father
- Sibling correlations in education and income are similar to those estimated in literature
- Correlations in education and income are substantially larger than in longevity

# Twin Correlations

| | Twins | | | |
|---|---|---|---|---|
| Outcome: | Lifespan | Education | HH Income | IGCLF |
| Brother/Brother | 0.206 | 0.738 | 0.488 | 0.106 |
| | (0.015) | (0.023) | (0.072) | (0.018) |
| N | 4,966 | 1,582 | 1,620 | 4,957 |
| Sister/Sister | 0.199 | 0.745 | 0.481 | 0.094 |
| | (0.018) | (0.026) | (0.041) | (0.021) |
| N | 4,334 | 1,094 | 1,107 | 4,320 |
| Sister/Brother | 0.053 | 0.571 | 0.412 | 0.107 |
| | (0.014) | (0.027) | (0.069) | (0.109) |
| N | 5,357 | 1,413 | 1,491 | 5,354 |

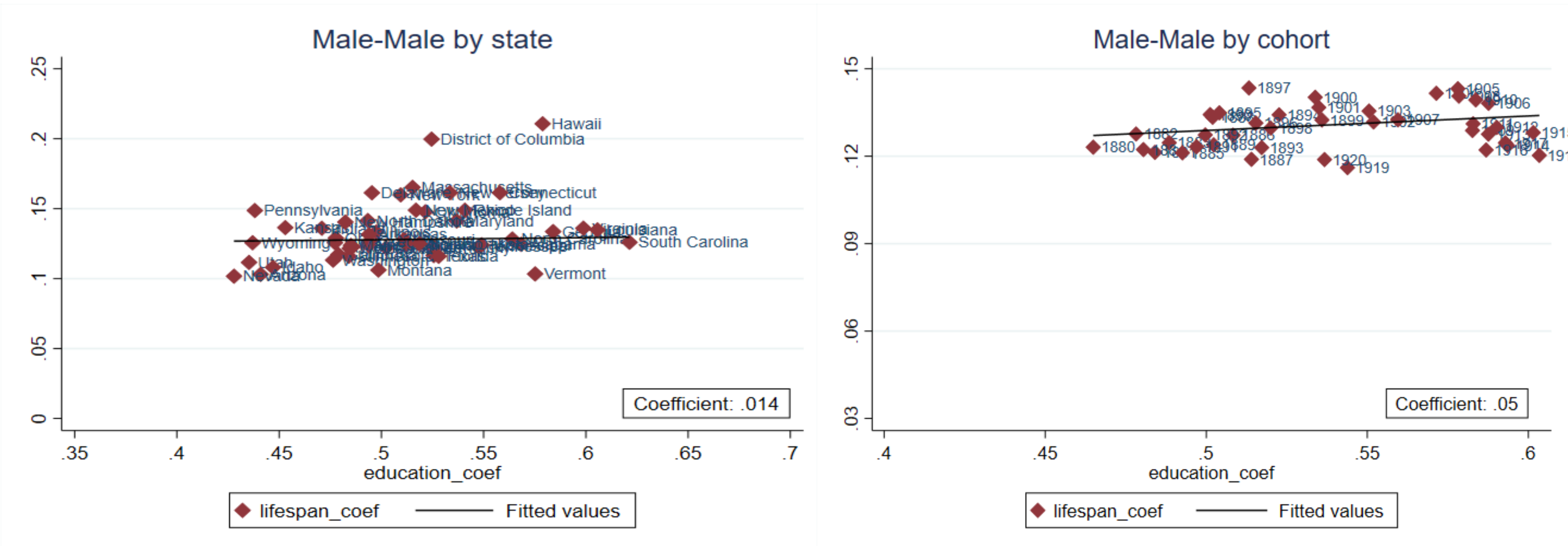
- The correlation in longevity is larger for twins than siblings. Similar results for education and income.
- Sister/brother twin correlations are similar to sister/brother sibling correlations
- IGCL with father is similar to full sample

# Role of SES in IGCL

| Sample | sample with education | | sample with education, income and occupation | | sample with education | | sample with education, income and occupation | |
|---|---|---|---|---|---|---|---|---|
| Parental Lifespan | Father | | | | Mother | | | |
| **Panel A: Son's lifespan** | | | | | | | | |
| Parental Lifespan | 0.0931 | 0.0918 | 0.0917 | 0.0904 | 0.0633 | 0.0607 | 0.0607 | 0.0606 |
| | (0.0010) | (0.0010) | (0.0010) | (0.0010) | (0.0008) | (0.0008) | (0.0009) | (0.0008) |
| Child's Education | | 0.2203 | 0.2148 | 0.2661 | | 0.2135 | 0.2111 | 0.2658 |
| | | (0.0036) | (0.0043) | (0.0041) | | (0.0037) | (0.0043) | (0.0041) |
| Income/100 | | | -0.0134 | 0.0015 | | | -0.0168 | -0.0008 |
| | | | (0.0019) | (0.0011) | | | (0.0021) | (0.0011) |
| Occupation | | | | -0.0501 | | | | -0.0535 |
| | | | | (0.0012) | | | | (0.0012) |
| N | 1,613,293 | | 1,456,619 | | 1,613,293 | | 1,456,619 | |
| $R^2$ | 0.0167 | 0.0191 | 0.0193 | 0.0205 | 0.0142 | 0.0165 | 0.0166 | 0.0180 |

- Very small effect on IGCL
- Similar results for daughters
- Education positive and significant, income and occupation of son negative (but positive for daughters)
- Weak test: more ideal test uses parental SES (not child)

# SES and Longevity Sibling Correlations

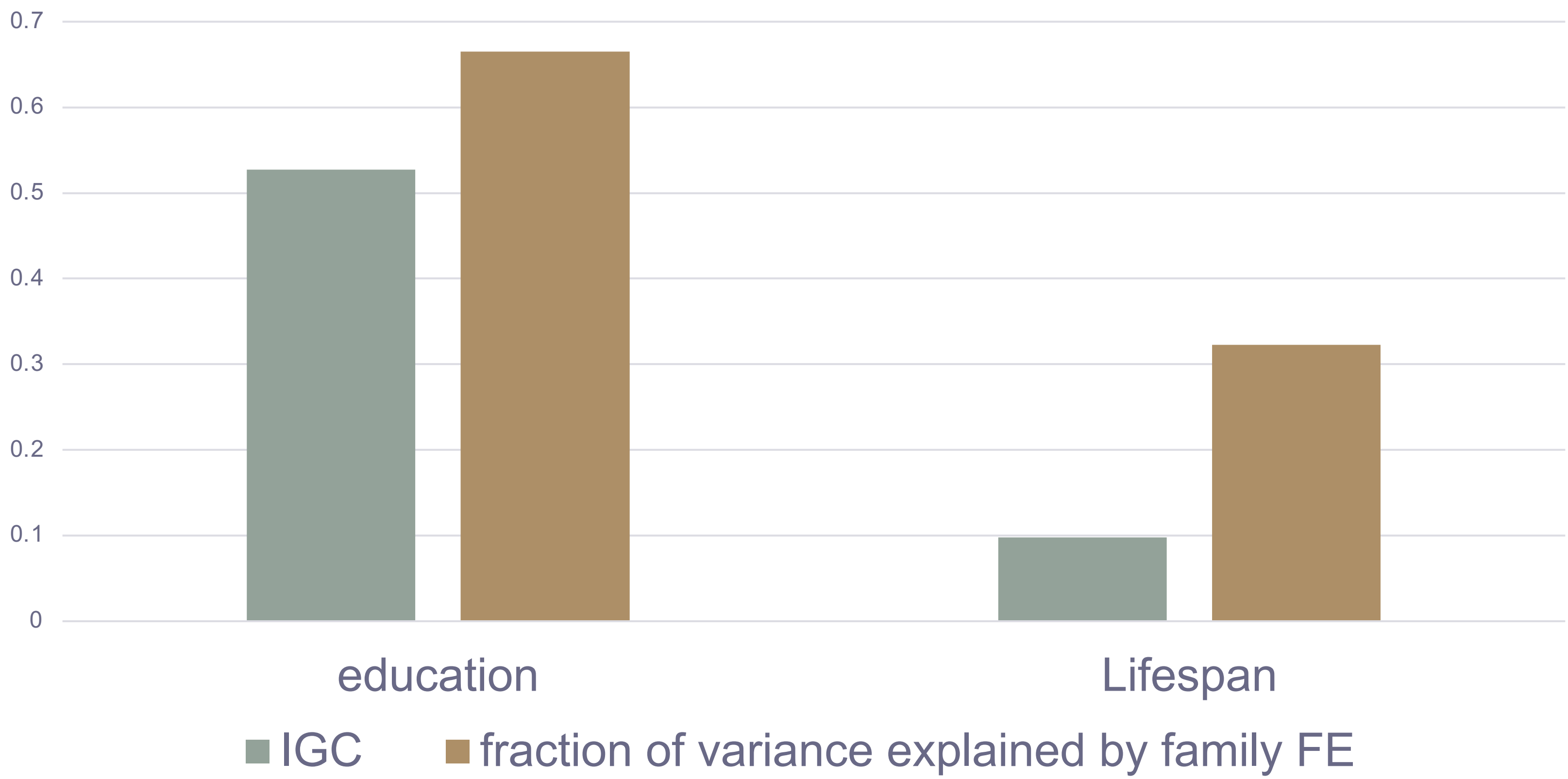



- Groups with higher correlations in education do not have higher correlations in longevity.
- Similar result for female-female estimates.

# Why is IGCL low?

1. Measurement error
    - IGCL remains low in states/places with good data
    - Robust to when restricted to individuals with death certificates in tree
2. Large variance in lifespan conditional on inherited traits
    - Vaupel (1988) demonstrates that even if the correlations between parents and children in inherited frailty is 1, IGCL can be low if the stochastic component of longevity is large.

# Variance Decomposition with Family FE




- Within family variance in longevity is much larger than for education

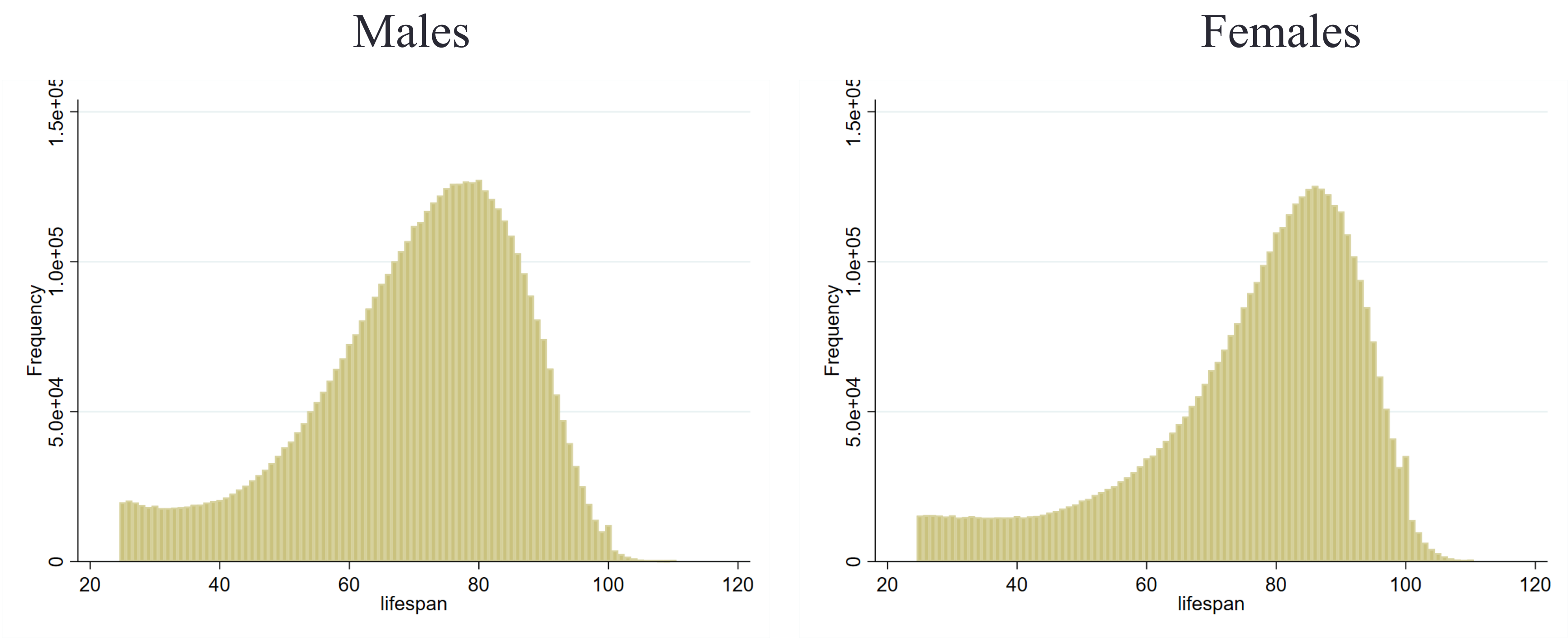
# Conclusion: Longevity Mobility Is High

- **IGCL is low**: ~ 0.1 regardless of the measure or controls
  - Similar to previous demographic work despite differences in population, time period, etc.

- **IGCL is much lower than correlations in SES** (education, occupation, income)
  - Lower than IGC in health measures ~0.2-0.3 (Halliday et al. 2017, 2020)
  - Low heritability because fixed family factors explain only a modest amount of the variation in lifespan

- **Simple additive model is a relatively good approximation**
  - But some evidence from sibling and grandparents correlations demonstrates the limits of this model

# Age at death distributions




Very little age heaping, except for age 100

[back](#)