# LIFE-M

**Longitudinal, Intergenerational Family Electronic Microdata**

PI: Martha Bailey (University of Michigan)

## Broad Aims

- Link individuals across 4 generations from birth to death certificates, including demographic (dob, race, place of birth), family (spouse/children), and economic information (occupation, education)
- Greatly expand the representation of women (crosswalk of birth and married names on birth certificates) and minorities

## Data

- Digitized vital records on births, marriages, and deaths (state
- 1880 and 1940 full count censuses

**Figure 1. Approximate Cohort and Generational Structure for Overview of Data Linking**

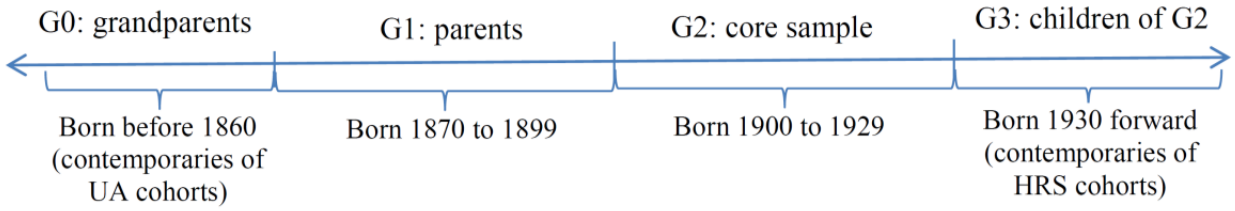
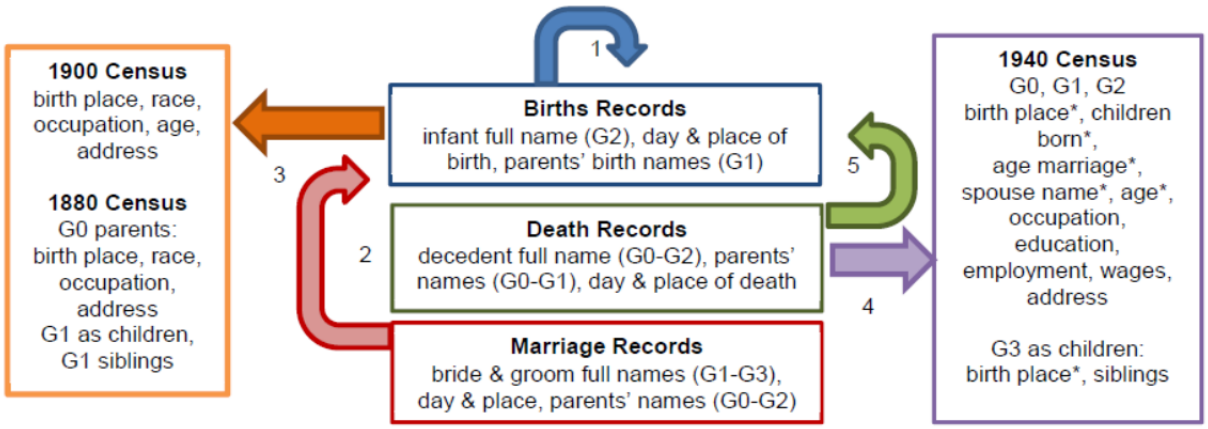



**Figure 2. Proposed Linkages of LIFE-M**




Key: G0: born <1860 (~UA cohorts); G1: born 1870-1899; G2: born 1900-1929; G3: born 1930- (~HRS cohorts).

## Schedule of Work

- Year 1: Create programs, processes to link core sample of OH births across all 5 datasets
  - All records "double trained" + discrepancy resolution process
- Years 2-3: Scale up to train much more primary data
- Year 4: Use machine learning model (together with trained data) to choose links for a much larger set of individuals; release data on ICPSR