

Creating the American Opportunity Study:

A NEW TOOL TO STUDY TRENDS IN SOCIAL MOBILITY

Trent Alexander (U.S. Census Bureau)

Jonathan Fisher (Stanford University)

David Grusky (Stanford University))

Carol House (National Academies - CNSTAT)

Michael Hout (New York University)

Amy O'Hara (U.S. Census Bureau)

Aliya Saperstein (Stanford University))

Timothy Smeeding (University of Wisconsin)

Matthew Snipp (Stanford University))

What data allow us to monitor social mobility?

- Longitudinal surveys help us focus on one cohort (e.g., NLSY) or a small sample (e.g. PSID)
- Last large-scale survey was in 1973! Occupational Changes in a Generation (OCG II) was a supplement to the CPS.
- Some linked administrative records (e.g., IRS and SIPP Gold Standard) tell us about income and earnings mobility for recent cohorts.

These data miss long-run TRENDS in social mobility

There have been massive changes that could lead to changes in social mobility.

- Family structure
- Gender roles
- Immigration
- Educational attainment
- Labor markets
- Economic inequality

“But the debate about social mobility and equality of opportunity is still unbalanced, with a surfeit of rhetoric and a deficit of data. The evidence base is too fragile for much of the political weight placed upon it. We need more data.” (Reeves, 2015)

Envisioning a New Tool

De facto intergenerational panel exists

- The decennial censuses
- Potential to add depth by linking survey and administrative data

Current efforts

North American Population Project

- Links historical (i.e., publicly available) decennial censuses (1850-1930 with 1940 on the way).

Census Bureau work

- Census Longitudinal Infrastructure Project (CLIP)
 - › Allows linkage of 1940, 2000, and 2010 Decennials, along with survey data such as ACS, CPS-ASEC, and SIPP.
 - › Can also bring in administrative data that can be linked.

The American Opportunity Study (AOS)

Goals

- Fill in the decennial holes between 1960 and 1990.
- Provide expertise to explore improved matching and linkage
- Assist in identifying key longitudinal research opportunities

A joint project between Stanford University, U.S. Census Bureau, and the National Academies of Sciences, Engineering, and Medicine.

The Hurdles

Name is not digitized in the 1950-1990 decennials.

- All other data from the decennials have been captured and digitized. But to link we need the person's name.
- **1990 forms** are stored on 120,000 reels of microfilm.
- Name is handwritten by the respondent.

Person #

Info

Last name,
First name,
Middle Initial

n capture

PLEASE FILL ONE COLUMN FOR EACH PERSON LISTED IN QUESTION 1a ON PAGE 1.

NAME PERSON 2 PERSON 3 PERSON 4 PERSON 5 PERSON 6

SON 2 PERSON 3 PERSON 4 PERSON 5 PERSON 6

1. **NAME**
Last name: *Rodriguez* First name: *Perry* Middle initial: *C*
Last name: *Ross* First name: *Ross* Middle initial: *J*
Last name: *Wilson* First name: *Jordan* Middle initial: *A*
Last name: *Smith* First name: *Wallace* Middle initial: *J*

2. **RELATIONSHIP**
If a RELATIVE of Person 1:
 Brother/sister
 Father/mother
 Grandchild
 Son/daughter
 Stepson/stepdaughter
 Other relative
 If NOT RELATED to Person 1:
 Roomer, boarder, or foster child
 Unmarried partner
 Housemate, roommate
 Other nonrelative

3. **SEX**
 Male Female

4. **RACE**
 White
 Black or Negro
 Indian (Amer.) (Print the name of the enrolled or principal tribe.)
 Eskimo
 Aleut
 Asian or Pacific Islander (API)
 Chinese
 Filipino
 Hawaiian
 Korean
 Vietnamese
 Other race (Print race)

5. **AGE**
 a. Age: 0 0 0 0 0 1 0 1 0 0 0 0
 b. Year of birth: 1 8 0 0 0 0 0 0 0 0 0 0

6. **MARRIAGE**
 Now married
 Widowed
 Divorced
 Separated
 Never married

7. **SPANISH/HISPANIC ORIGIN**
 No (not Spanish/Hispanic)
 Yes, Mexican, Mexican-Am., Chicano
 Yes, Puerto Rican
 Yes, Cuban
 Yes, other Spanish/Hispanic (Print one group, for example: Argentinian, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.)

8. **FOR CENSUS USE**
 2822 503 1815

A pilot test



Where we are now

Rented two microfilm scanners from different companies.

- 608 reels scanned (197 long form; 411 short form)
 - Some overlap between the two scanners
- Also had National Archives scan two reels
- Have ~1 million images (i.e. households)

Hand-keyed ~25,000 names - some double-keyed

- ~95% exact match rate on double-keyed images

Where we are now

First optical character recognition (OCR) vendor digitized names last week. Second OCR vendor is working as we speak.

Names will then be assigned a **Protected Identification Key (PIK)**, which allows for linkage to other surveys.

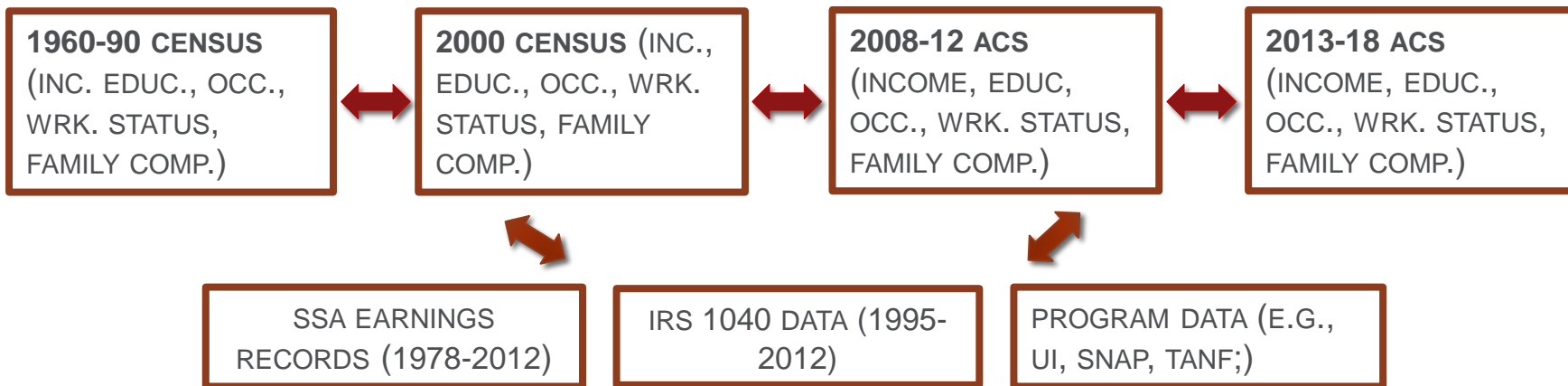
Also found new data!

- Master Address File IDs for all respondents.
 - Allows for address linkage using 1989 tax returns.
- Last name and first initial for ~27 million household heads living in rural routes or multi-unit buildings.
 - Allows for identifying head of household and potentially spouse using tax records and 2000 Decennial.
 - Have way to identify kids as well if they were still in the house in 1995 and/or 2000.

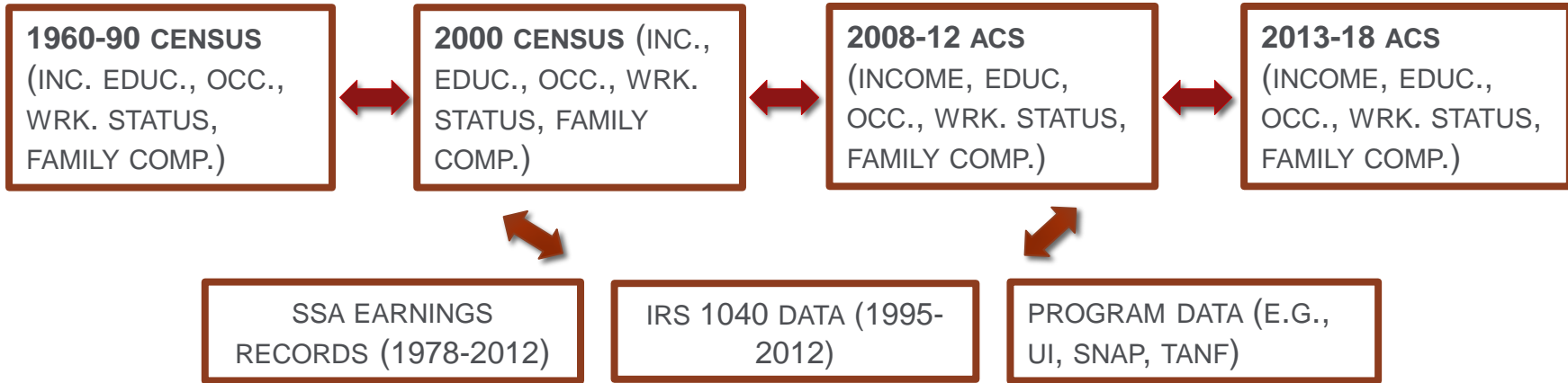
STEP #1: LINKING RECORDS FROM ACS AND CENSUS ACROSS MANY YEARS



STEP #2: ADDING IN ADMINISTRATIVE DATA



STEP #3: LINKING CHILDREN WITH PARENTS

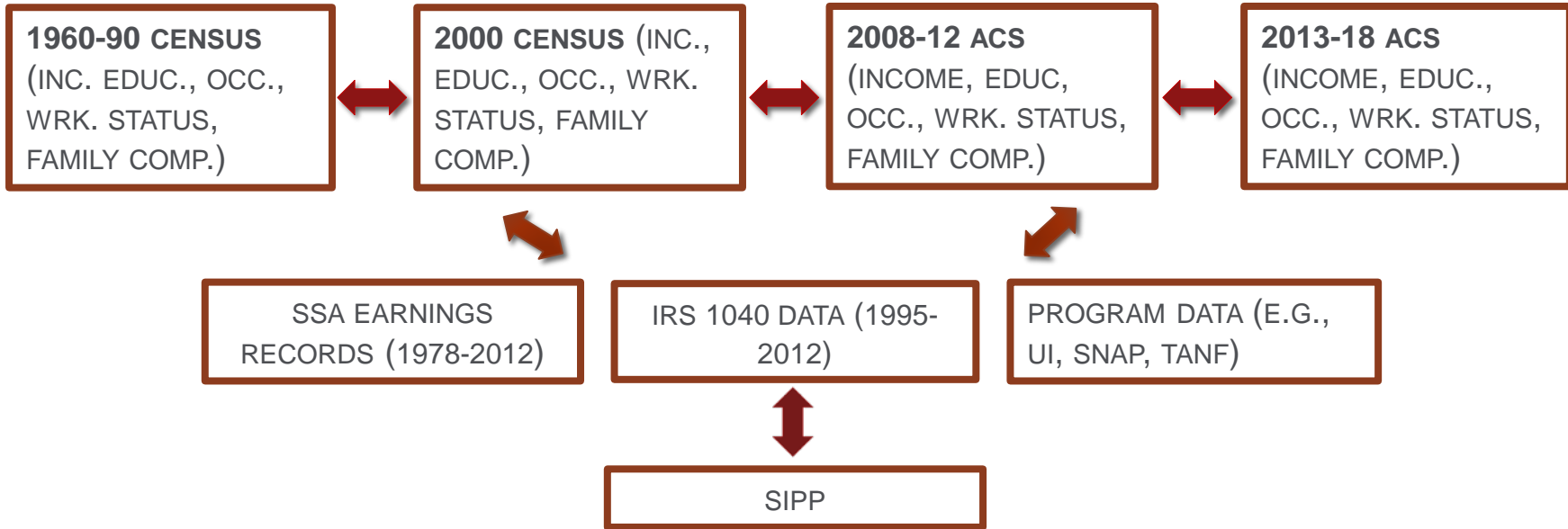


KIDLINK FILES (E.G., FORM SS-5 IF HISTORICAL FORMS AVAILABLE)

CORESIDENCY IN ACS, CENSUS LONG/SHORT FORM

PARENTAL REPORTS OF CHILDREN'S SSN TO IRS SINCE MID 1980'S

STEP #4: SLIPPING IN THE SURVEY



SURVEYS WITH IDENTIFIERS CAN BE SLIPPED IN (E.G., SIPP)

THE SURVEY NOW AS A LEAN AND MEAN VALUE-ADDED INSTRUMENT

Access will be through Census Bureau and its network of Research Data Centers

- AOS focuses on social mobility.
- But when 1990 Decennial can be linked, it will also be part of CLIP, which means program analysis and other research can be conducted with the linked data.