

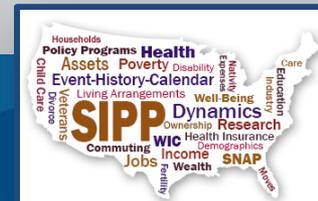
Content: SIPP Variables

Person Level Information

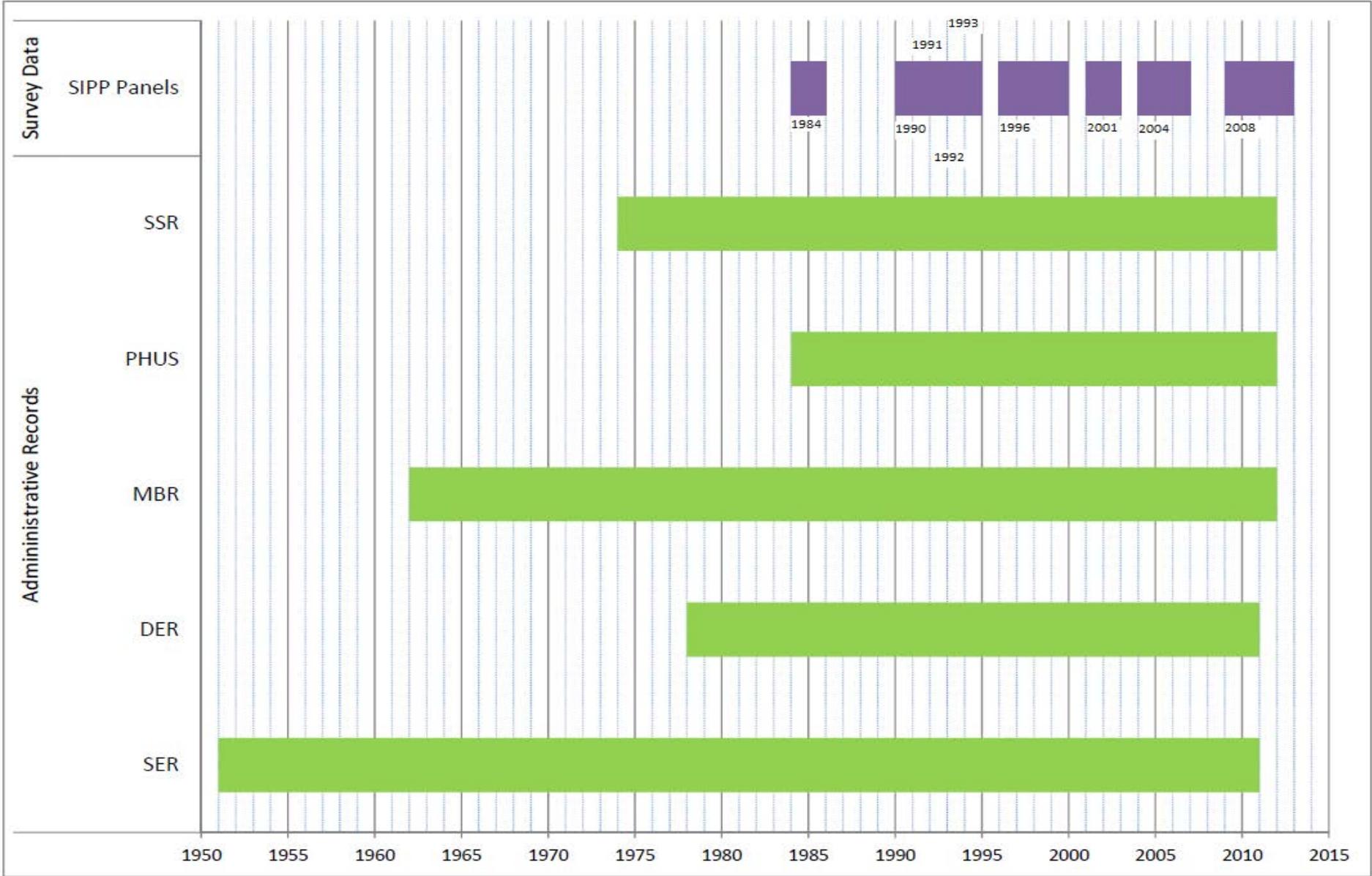
- Marital history for up to 4 marriages
- Fertility history
 - Number of own kids (for men and women) * First birth year and last birth year for biological children (women)
- Nativity / Time period arrived in the U.S.
- Race / Ethnicity (Hispanic vs. non-Hispanic)
- Education
 - Level * Field of Bachelor's degree * Years started high school and post-high school education * Years received high school diploma & bachelor's degree * Current enrollment in college or high school
- Wealth (Home ownership, Home equity, Non-housing wealth)
- Enrollment in pension (Defined benefit * Defined contribution)
- Industry & Occupation
 - 4 category industry (manufacturing, wholesale/retail, FIRE/public admin/military, agriculture/mining/construction)
 - 3 category occupation (managerial/professional, technical/sales, other)
- Disability (Work-limiting * Work-preventing)

Person-Month Level Information

- Income, Earnings, Hours and Weeks worked
- Layoff indicators
- Health insurance and if Employer-provided
- Workers & Veterans compensation (Receipt & Amount)
- Program Participation
 - AFDC/TANF * Food stamps



Design and Connection of Data Sources

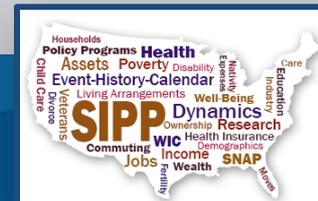


Some Possible Topics

- Wage Inequality
- Retirement patterns
- Disability applications
- Immigrant outcomes
- Education outcomes
- Fertility history questions
- Spouse behavior

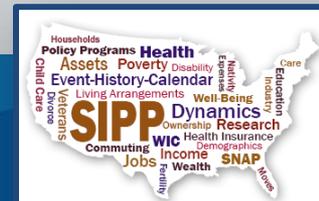
Ongoing SSB Research

- Persistence of breadwinning across couples
- Associations with graduating during a recession
- Earnings and employment trajectories across demographic groups
- Labor market outcomes for disabled workers
- Wage mobility among public assistance recipients



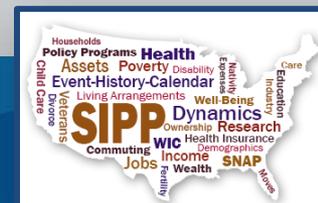
Topic Model Imputation

- How to implement new imputation methods and still release data in a timely manner for a survey with 11,000 variables?
 - Replace item-level hot deck with parametric model-based approach
 - Helps handle small stratifying cell size problem
 - Allows inclusion of many more predictor variable
 - SIPP SSB provides the methodological foundation for modelling
 - Use administrative data to mitigate problems caused when survey data are not “missing at random”
- Solution
 - Use topic flags as alternative to whole-record donation for cases where respondent did not complete the majority of the survey.
 - Indicator variables for all the major topics covered by SIPP
 - Implement new methods only for these 40+ variables



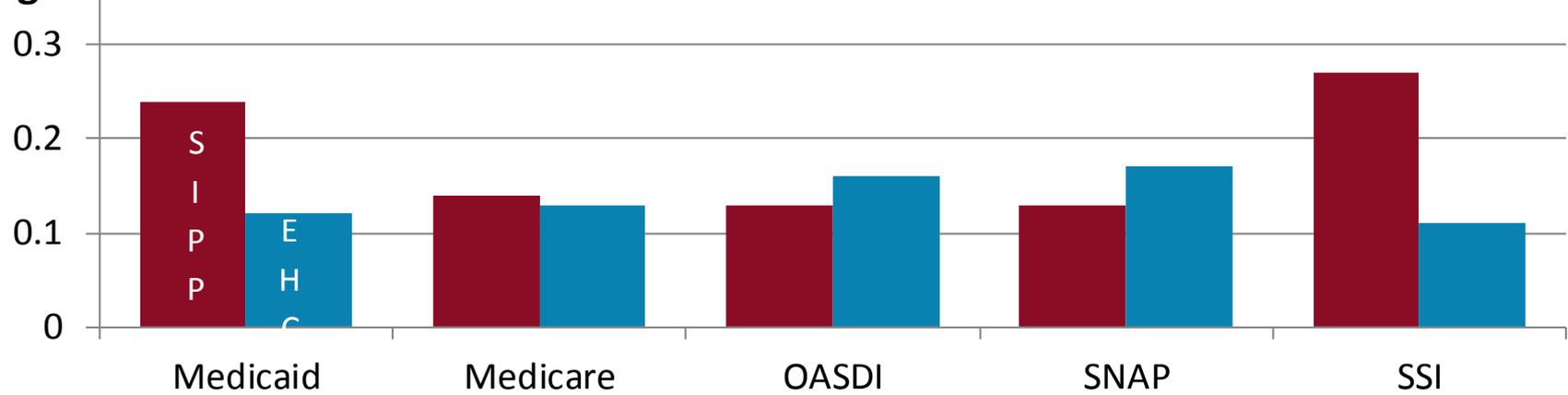
Purpose of topic flags

- Measure number of missing topics
- Facilitate imputation of missing data
 - Stop whole-person substitution
 - Preserve correlation across topics by estimating a joint distribution for imputation
 - Allow any reported data to be used, including from other family members
 - Use administrative data as additional predictors
- Use in downstream edits:
 - Topic flag sets the universe for follow-up questions
 - Flags from other topics can be used in edits and hot decks

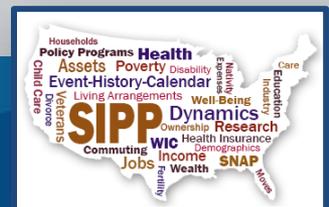
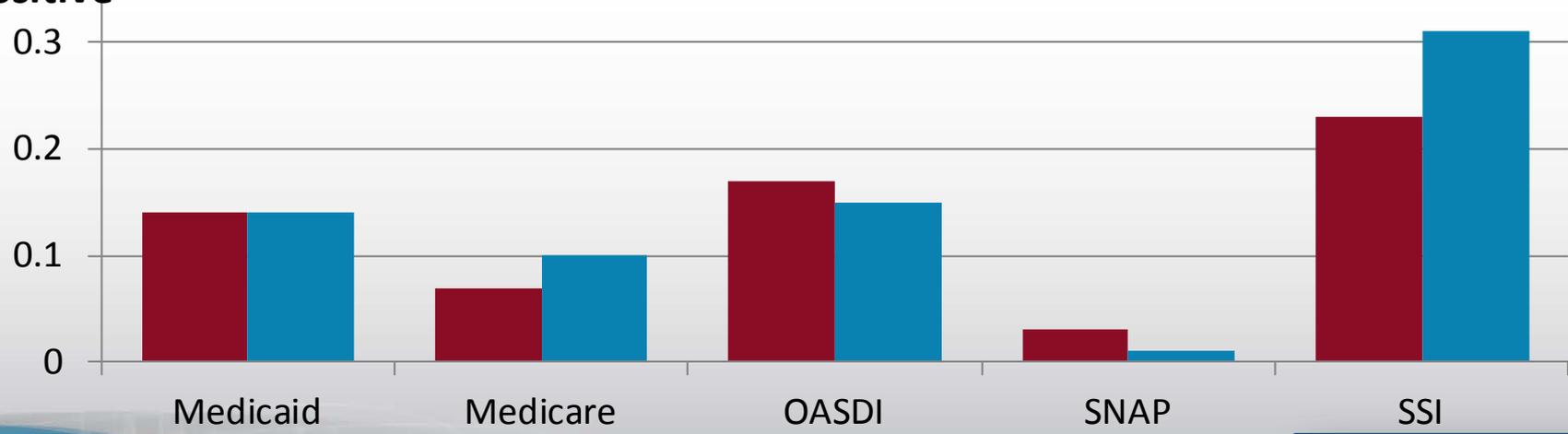


Comparing **SIPP** and **SIPP-EHC** With Linked Administrative Data

False Negative

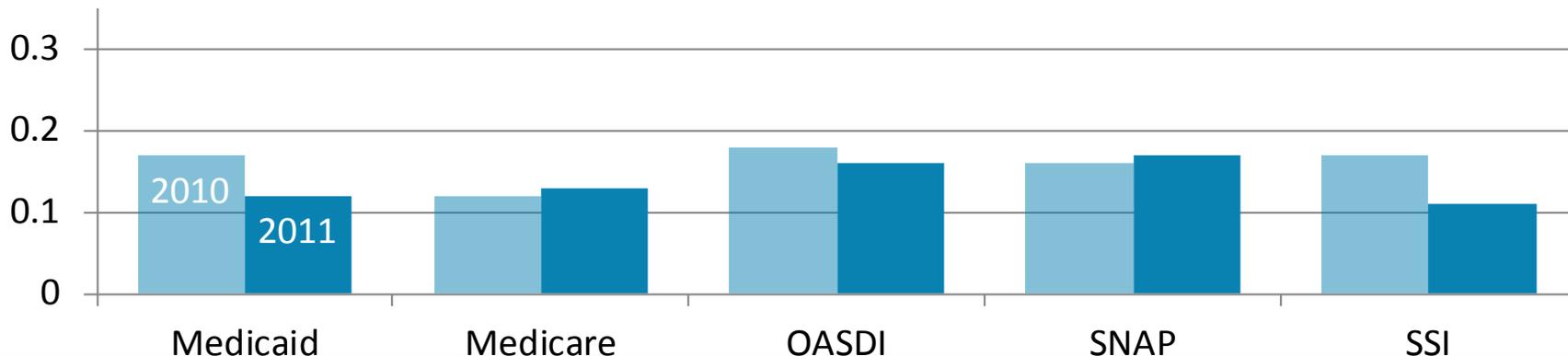


False Positive



Comparing SIPP-EHC From CY2010 and CY2011 With Linked Administrative Data

False Negative



False Positive

