

Producing U.S. Population Statistics Using Multiple Administrative Sources

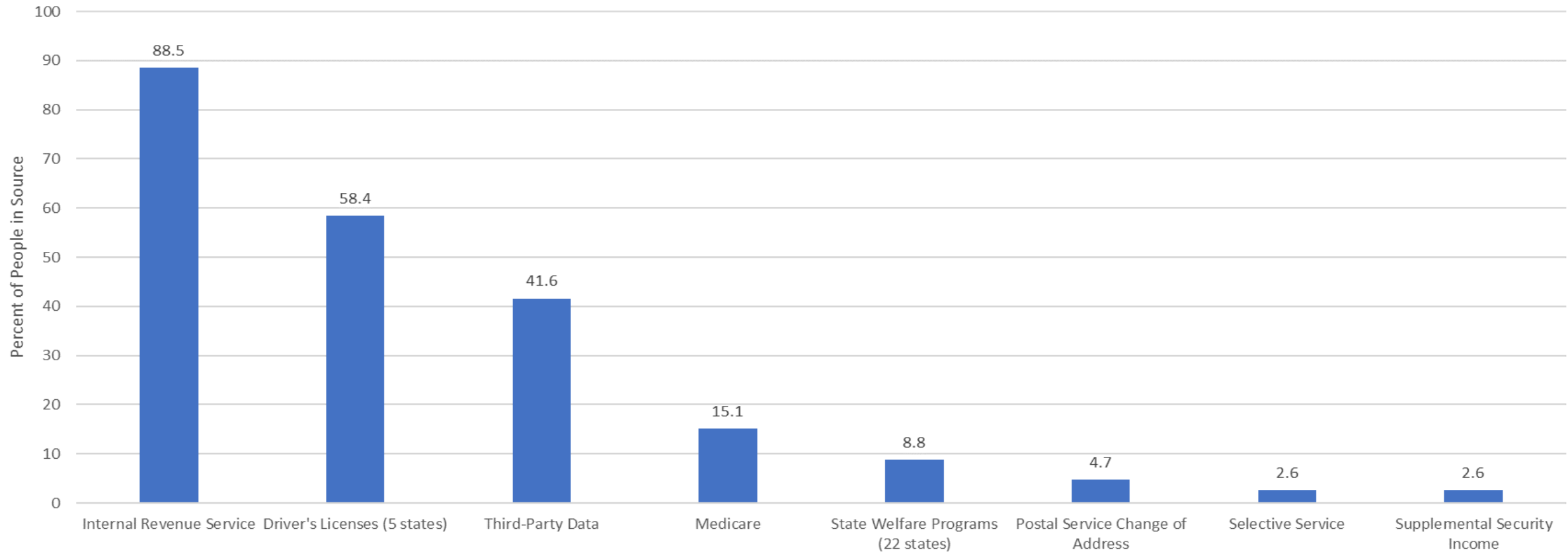
J. David Brown and Marta Murray-Close

December 6, 2023

Challenges in Constructing Administrative Record-Based Population Statistics

- Covering all eligible people
- Excluding ineligible people
- Coverage consistency
- Locational accuracy
- Demographic characteristic consistency

Percent of AR Census People in Source



States providing driver's license data include Alabama, Iowa, Nebraska, South Carolina, and South Dakota. Driver's license data raised AR census population estimates by 3.0 percent in this group of states. State welfare program data raised AR census population estimates by 2.0 percent in the states providing those data.

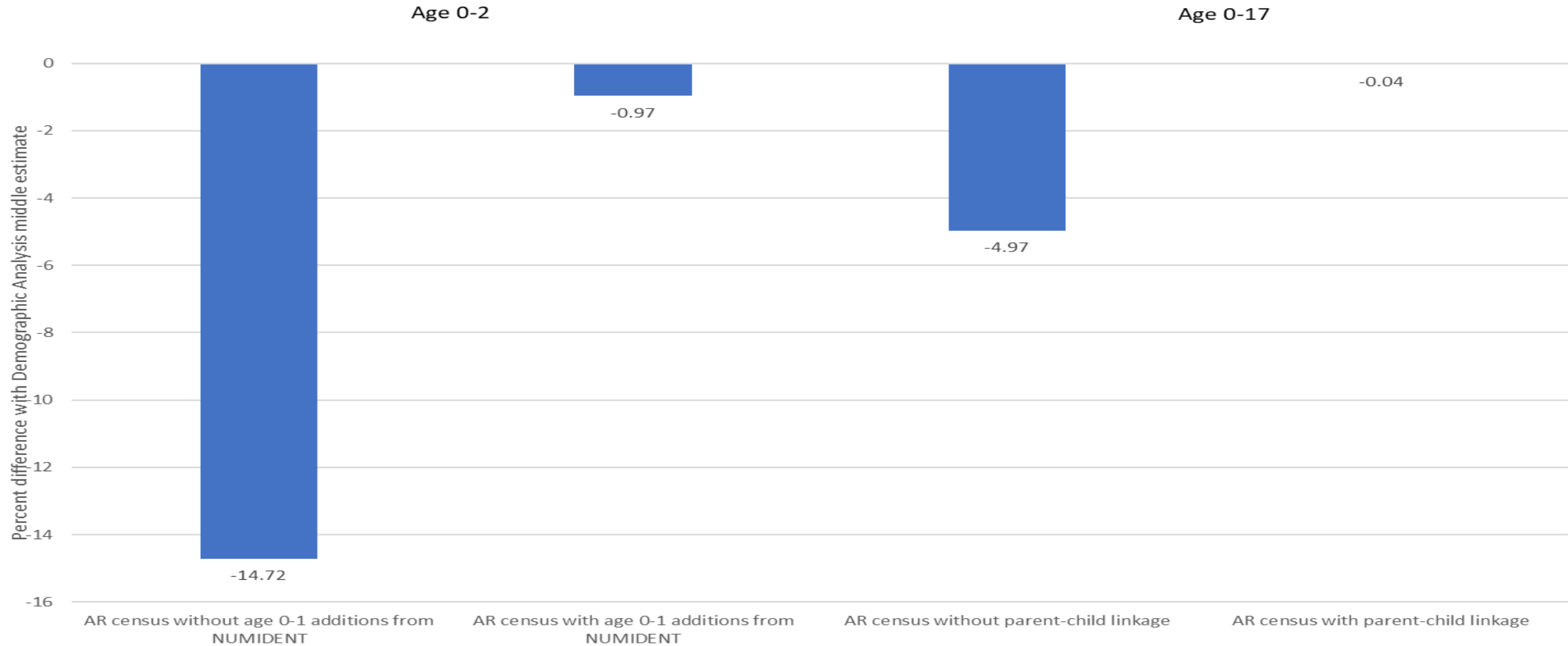
Coverage of Children

- Most AR sources do not cover children
- Main source is individual income tax forms
 - Not all people file taxes, and infants appear in forms with lag
- U.S.-born children have Social Security number (SSN)
 - Social Security Administration data contain birth city, not current residence
- We link children to their parents
 - Children may not live with their parents

Demographic Analysis Benchmark

- Uses birth and death records for native-born population
- Three sets of estimates using different assumptions
- Thought to be most reliable estimates for children

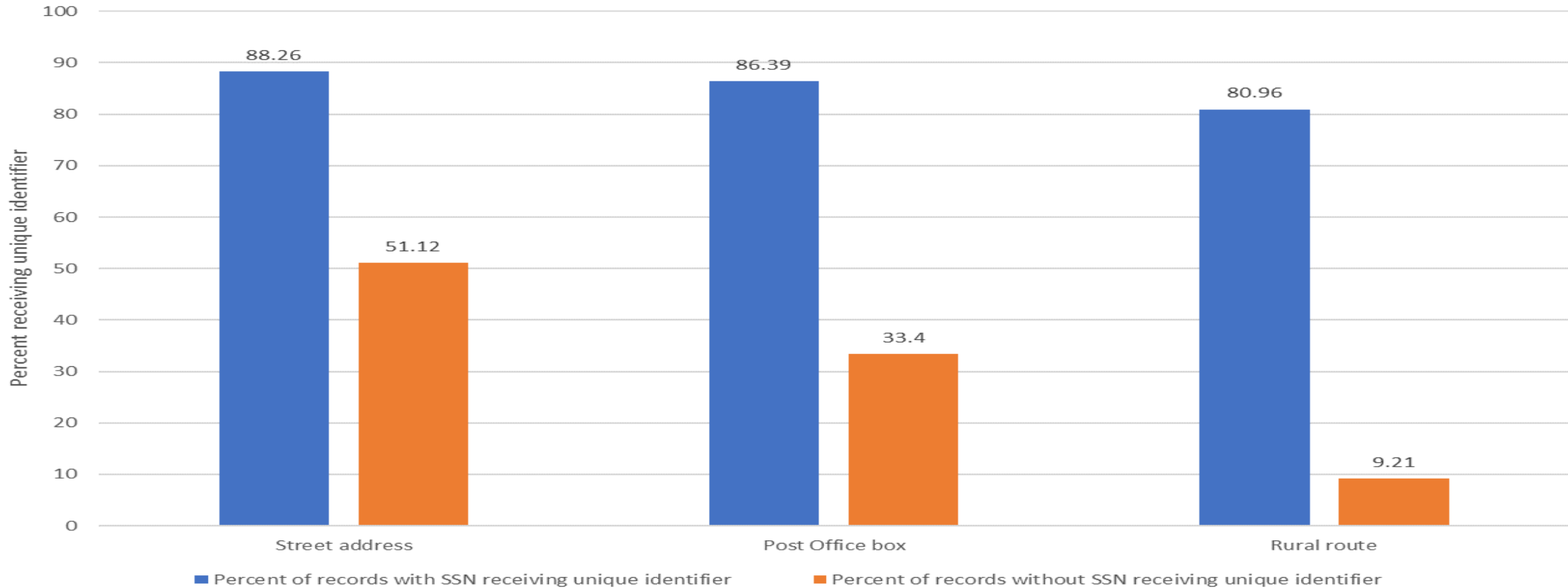
Effects of Methods to Increase Child Coverage



Record Linkage

- We use only records that can be assigned unique person identifier
 - Ensures that people are counted just once
 - Used to attach demographic characteristics and location
- Most reliable record linkage uses SSN
- Some AR lack SSNs
- Linking using address is next-best solution
- Linkage using Post Office (P.O.) Box or rural route is less effective than street address

Effects of Address Type on Record Linkage



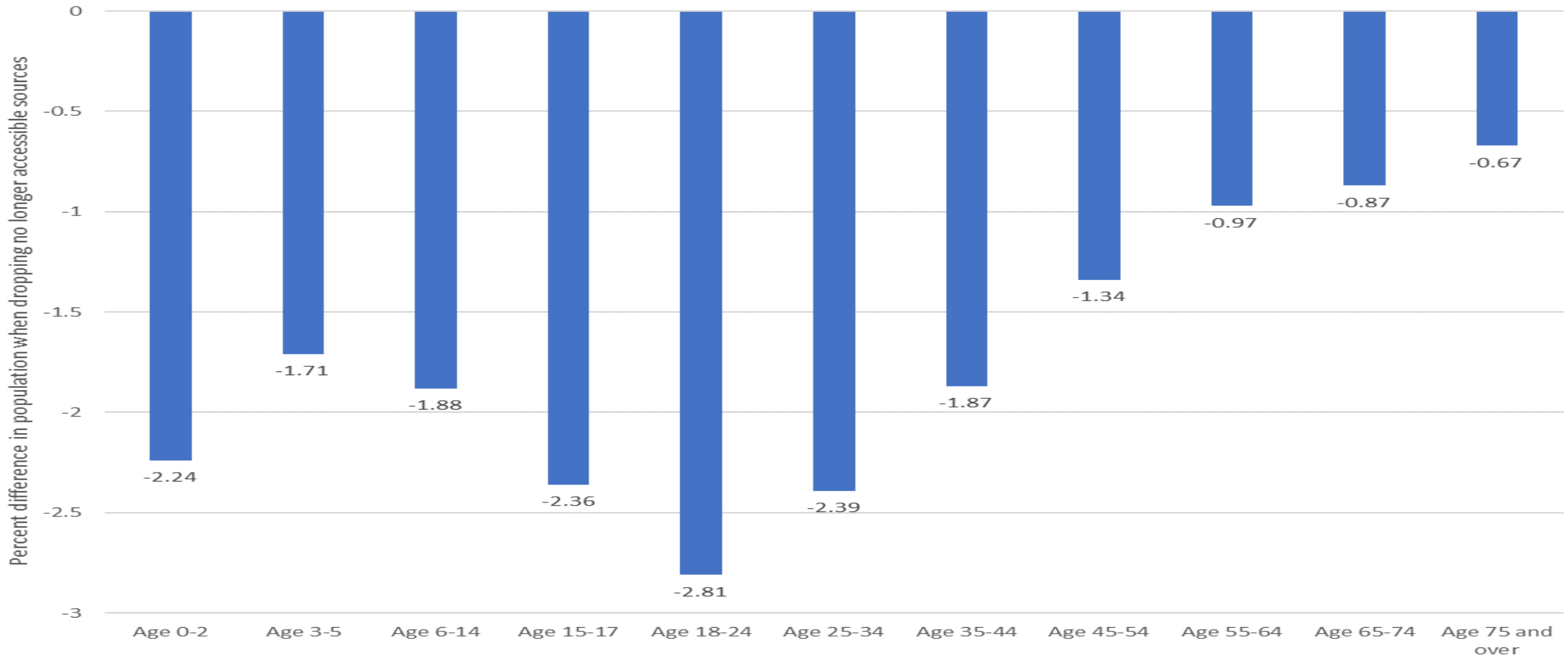
Excluding Ineligible People

- Exclude people not alive on reference date
- Use records with U.S. address and vintage near reference date
- Will try to develop model predicting whether people not found in recent AR data are still U.S. residents

Coverage Consistency

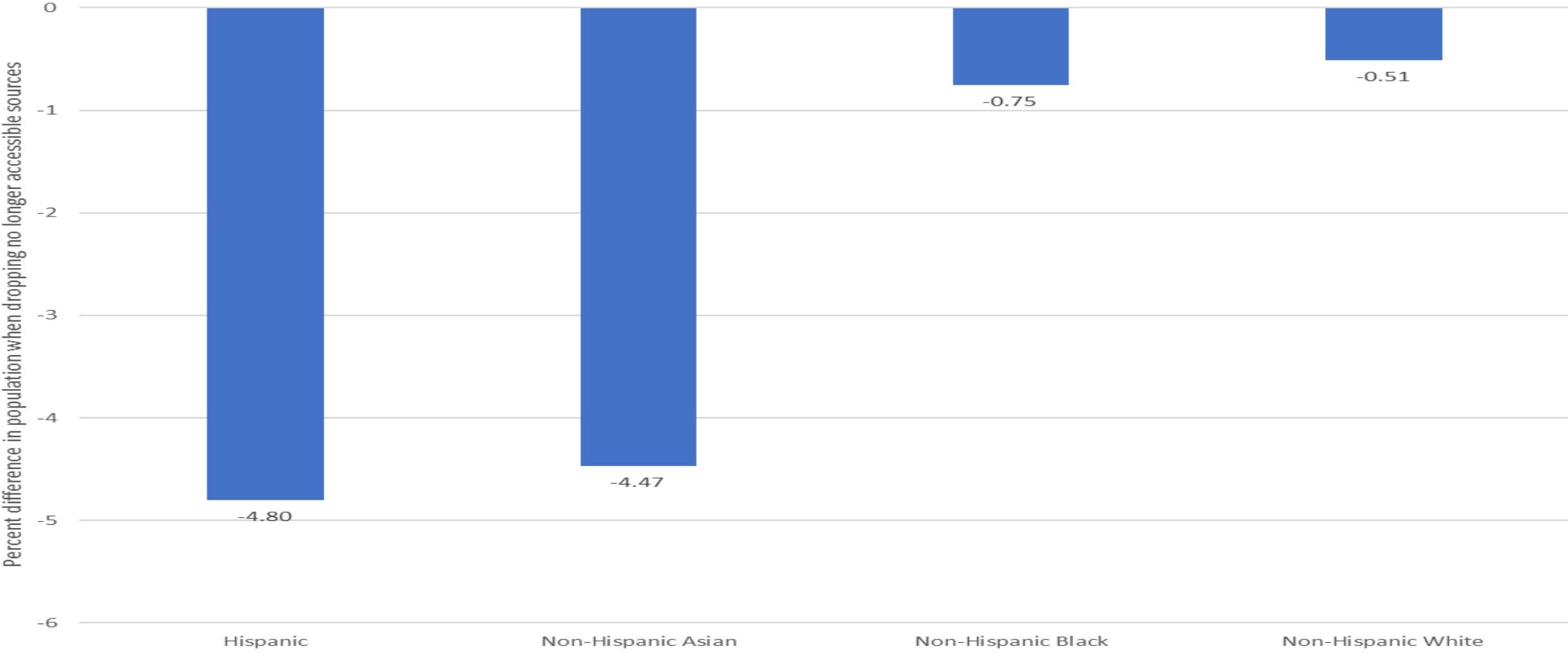
- Access to same sources over time
- Timely receipt of sources
- Consistent coverage by source over time
- People in only one source have less consistency

Effects of Loss of Source Access by Age Group



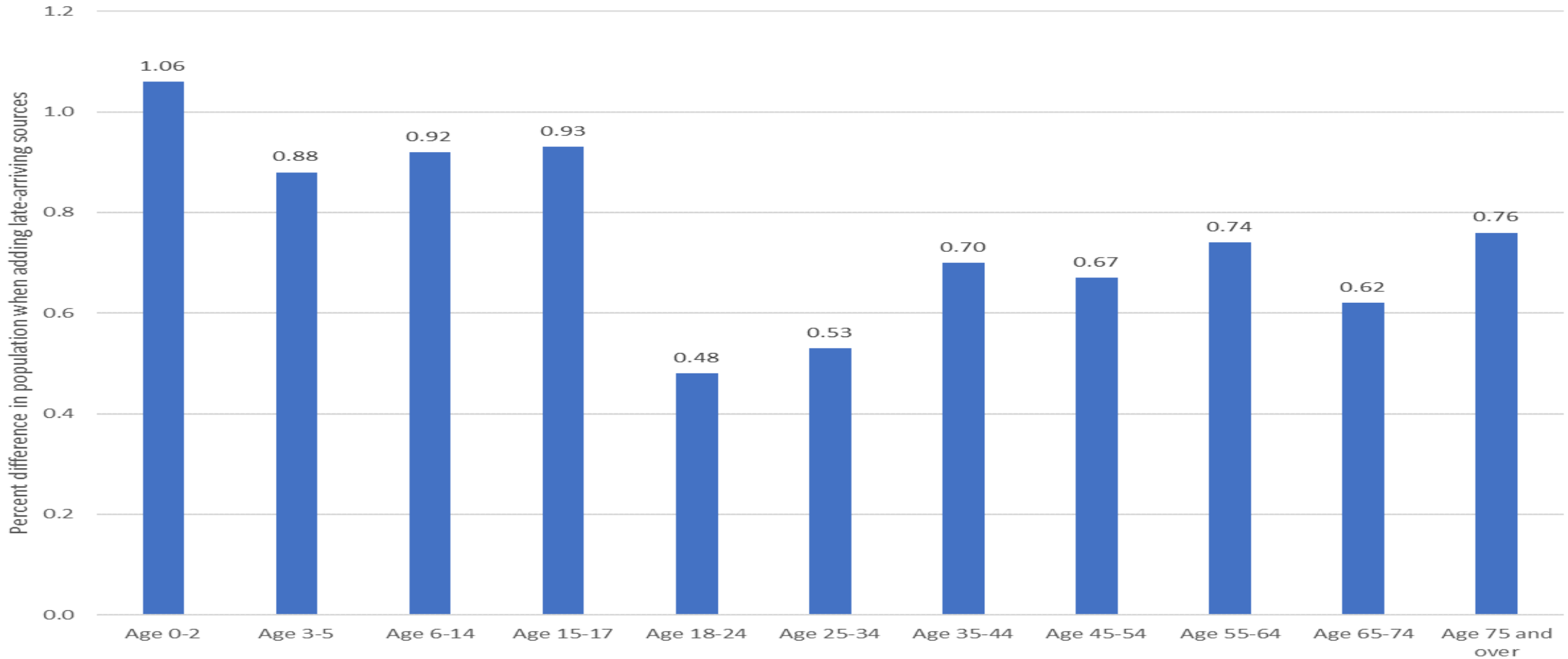
Lost sources focus mainly on noncitizens.

Effects of Loss of Source Access by Race/Ethnicity



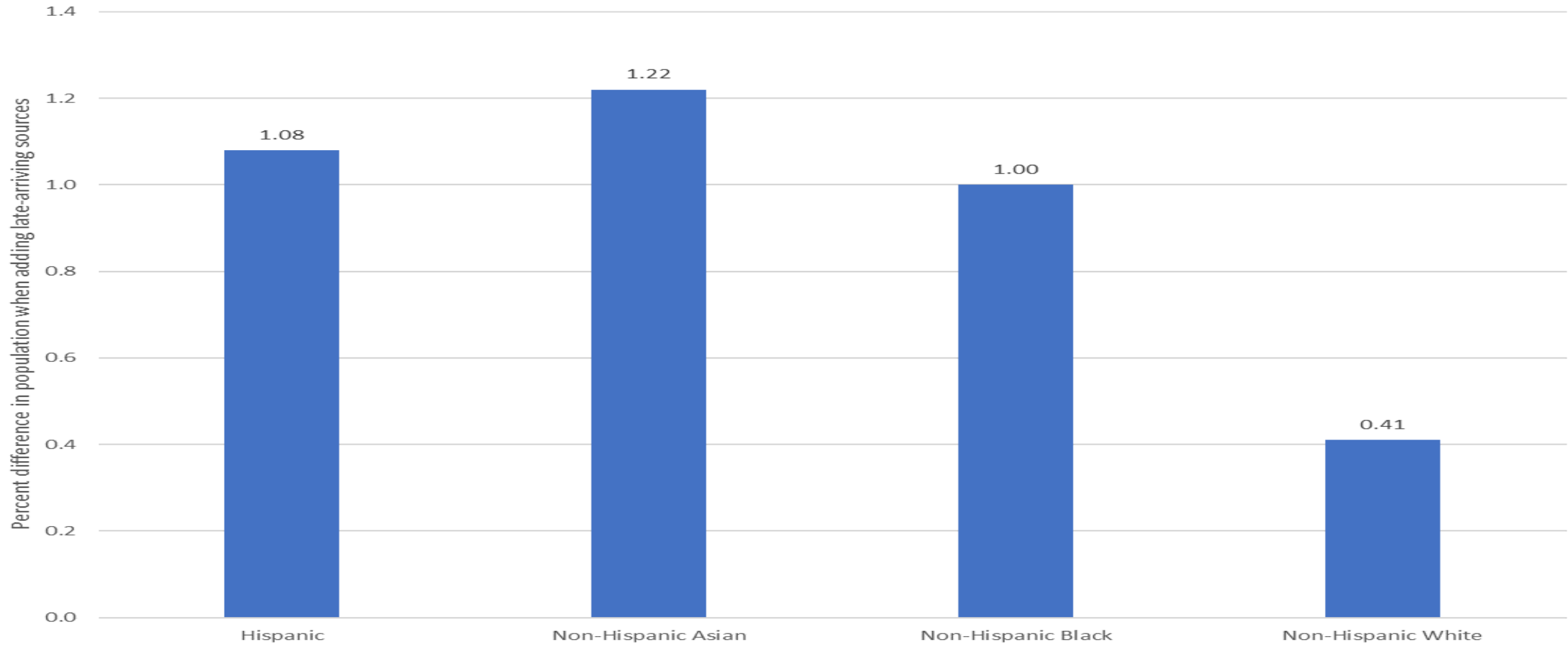
The data presented in this figure are approved for dissemination by the DRB (CBDRB-FY23-0253).

Effects of Adding Late-Arriving Sources by Age Group

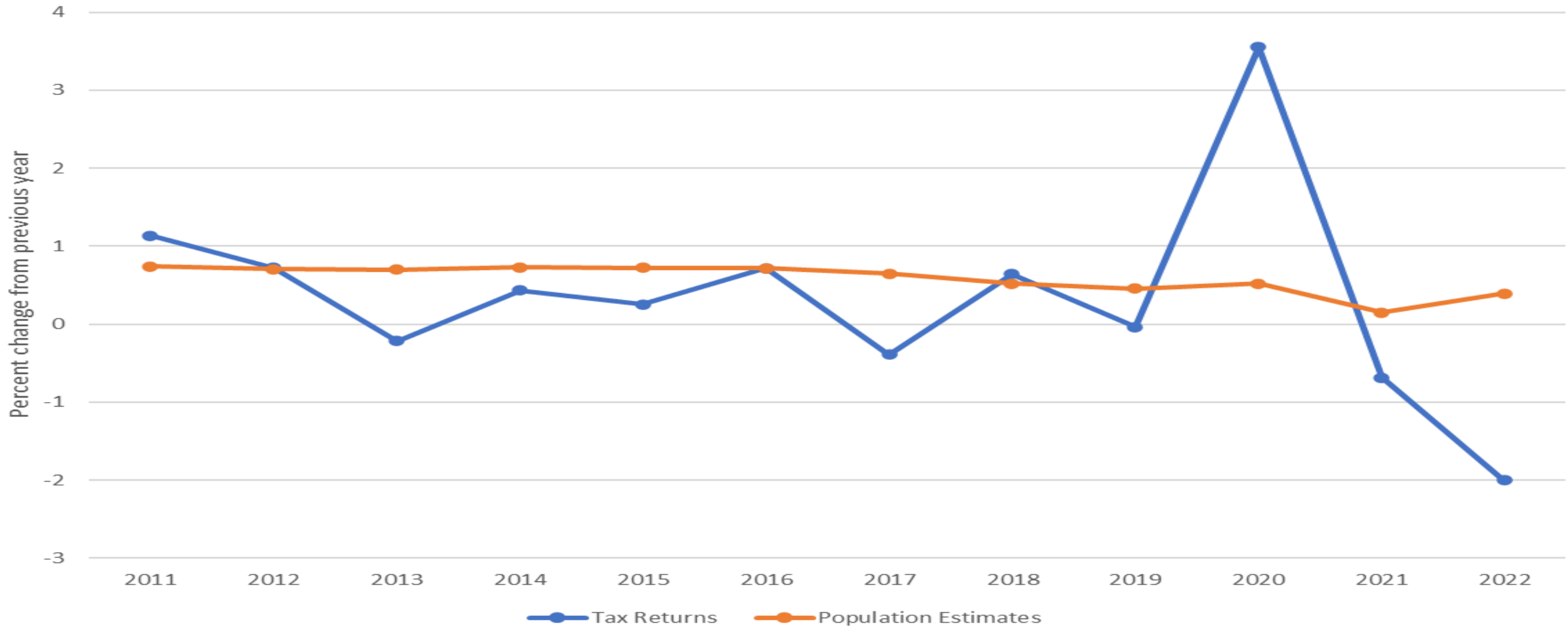


Medicaid and some tax returns arrived late.

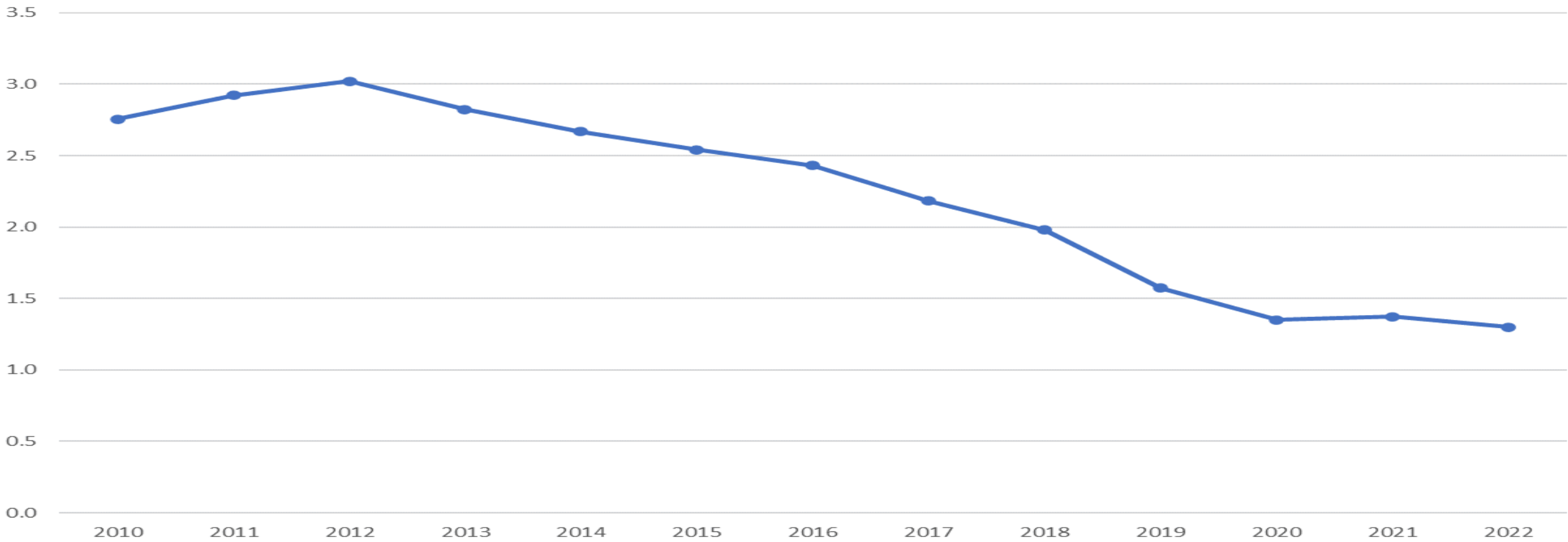
Effects of Adding Late-Arriving Sources by Race/Ethnicity



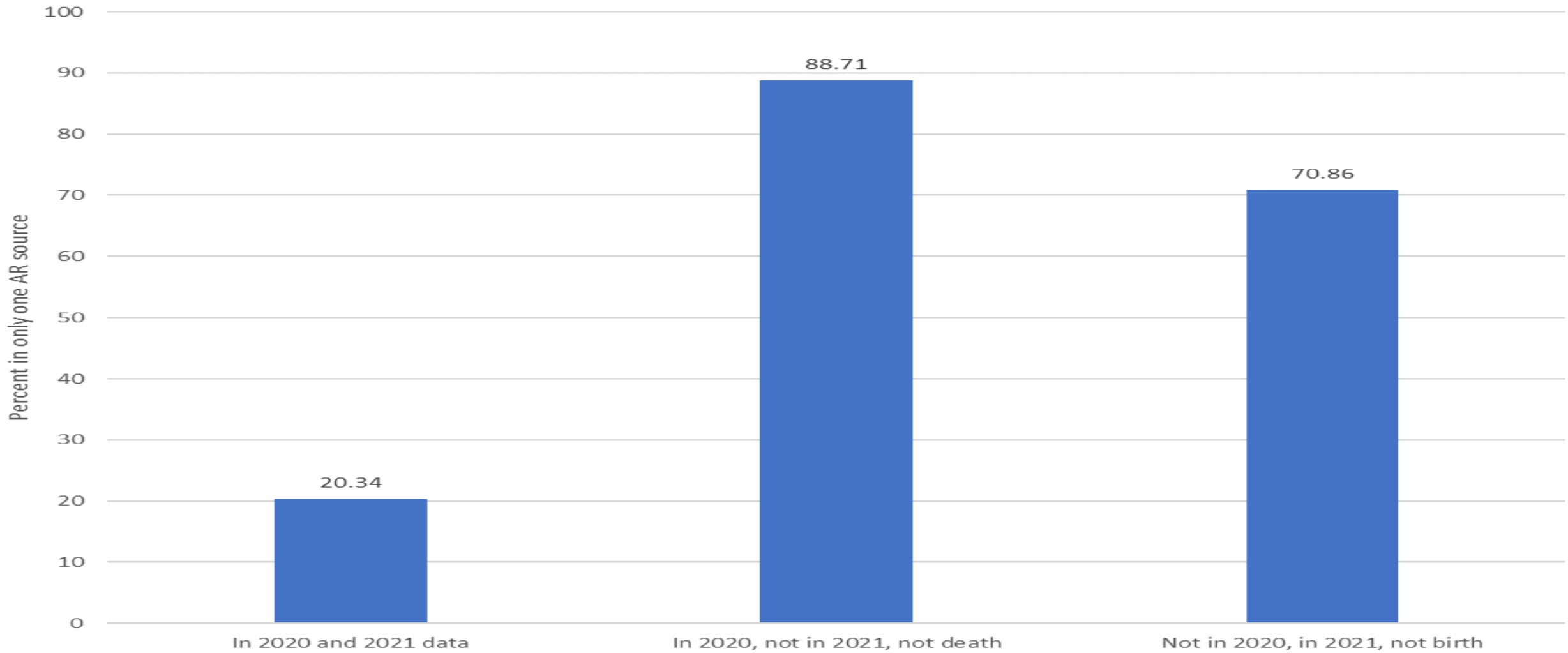
Percent Change in People in Tax Returns and the Population



Percent of Population with ITINs



Most People Entering or Exiting AR-Based Population Statistics are in Only One Source



Locational Accuracy

- 2.6% of people in AR census lack subcounty geography
- Impute lower-level geography for people without it
- 45.8% of people in AR census have more than one address
- Estimate person-place model to predict probability that address is the person's residence on reference date
 - Weight addresses by this probability

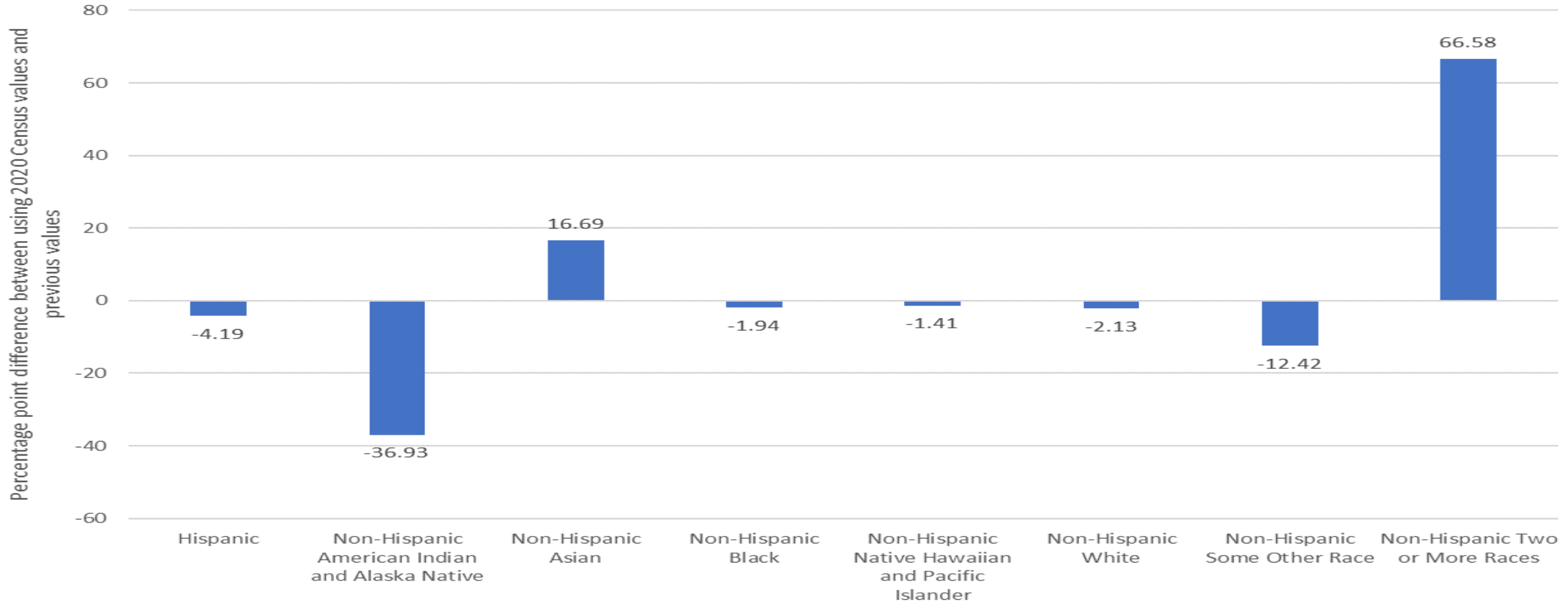
Demographic Characteristic Consistency

- Race/ethnicity questions changed in 2020, likely to change again by 2030
- AR sources may use different race/ethnicity definitions or not collect it at all
- We will try to develop model to adjust historical race/ethnicity responses for changes in race/ethnicity questions over time

AR Census Using Past Responses Compared to 2020 Census Responses

- AR census uses past responses for race/ethnicity
- In this comparison, we replace the past responses with 2020 Census responses for AR census people who can be linked to the 2020 Census
- This allows us to focus on effects of differences in reporting by the same people

Differences in Race/Ethnicity Distribution with 2020 Census Values Relative to Earlier Values



Conclusion

- Benefits of using multiple sources in AR-based population estimates
 - More comprehensive coverage
 - More consistent coverage
 - Improves prediction about where a person lives
- Difficulties when using multiple sources
 - Must address discrepancies across sources
 - More complex and time-consuming processing