

Using Tax and Social Security Records to Estimate Foreign-Born Emigration

Presented at the Annual Meeting of the
Population Association of America
New Orleans, LA
April 12-15, 2023

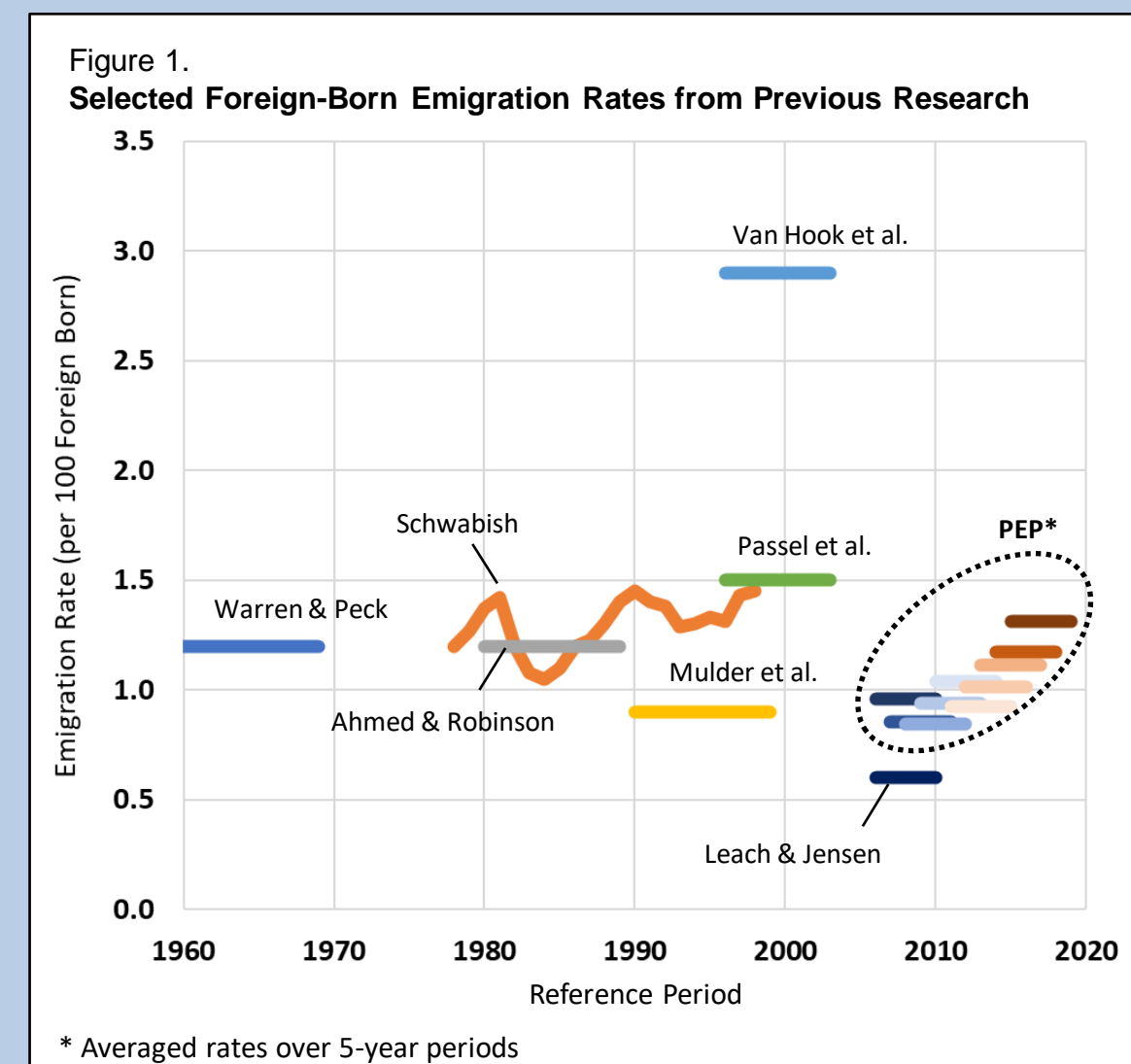
Thomas B. Foster and Erik Vickstrom, Center for Economic Studies, U.S. Census Bureau
Anthony Knapp, Population Division, U.S. Census Bureau

INTRODUCTION

Emigration from the United States is an important component of population change, but measuring it directly is difficult and current methods of estimation rely on dated and often unreliable data sources. This research takes the initial steps towards an estimate of foreign-born emigration derived from linked individual-level data from the Internal Revenue Service (IRS), Social Security Administration (SSA), and a composite administrative record file. We use these linked data to estimate foreign-born emigration rates for the 2010 to 2019 period at the national level, as well as by age and select countries of birth. Preliminary results are promising and show that emigration estimates derived from linked IRS-SSA records are comparable to those currently produced by the U.S. Census Bureau's Population Estimates Program, both overall and with respect to the distribution of emigrants across various individual characteristics.

BACKGROUND

Given the incompleteness of federal data on emigration, the Census Bureau and other researchers rely on indirect estimates of emigration based on population data from decennial censuses up to 2000 and the annual American Community Survey (ACS) since 2006. The **Population Estimates Program (PEP)** uses an ACS-based residual method. However, these methods are susceptible to changes in coverage and survey errors, which may yield volatile rates. Schwabish (2009) used longitudinal administrative data, which potentially obviates some of the limitations of survey data. Our research investigates this further.



DATA AND METHODS

For this research, we adopt a **signs-of-life** framework to estimate foreign-born emigration from the U.S. Under this framework, we first construct a starting population from administrative microdata and then track this population over tax years 2009 to 2020. We look for evidence that a person is still living in the U.S. (i.e., signs of life) across individually-linked data sources:

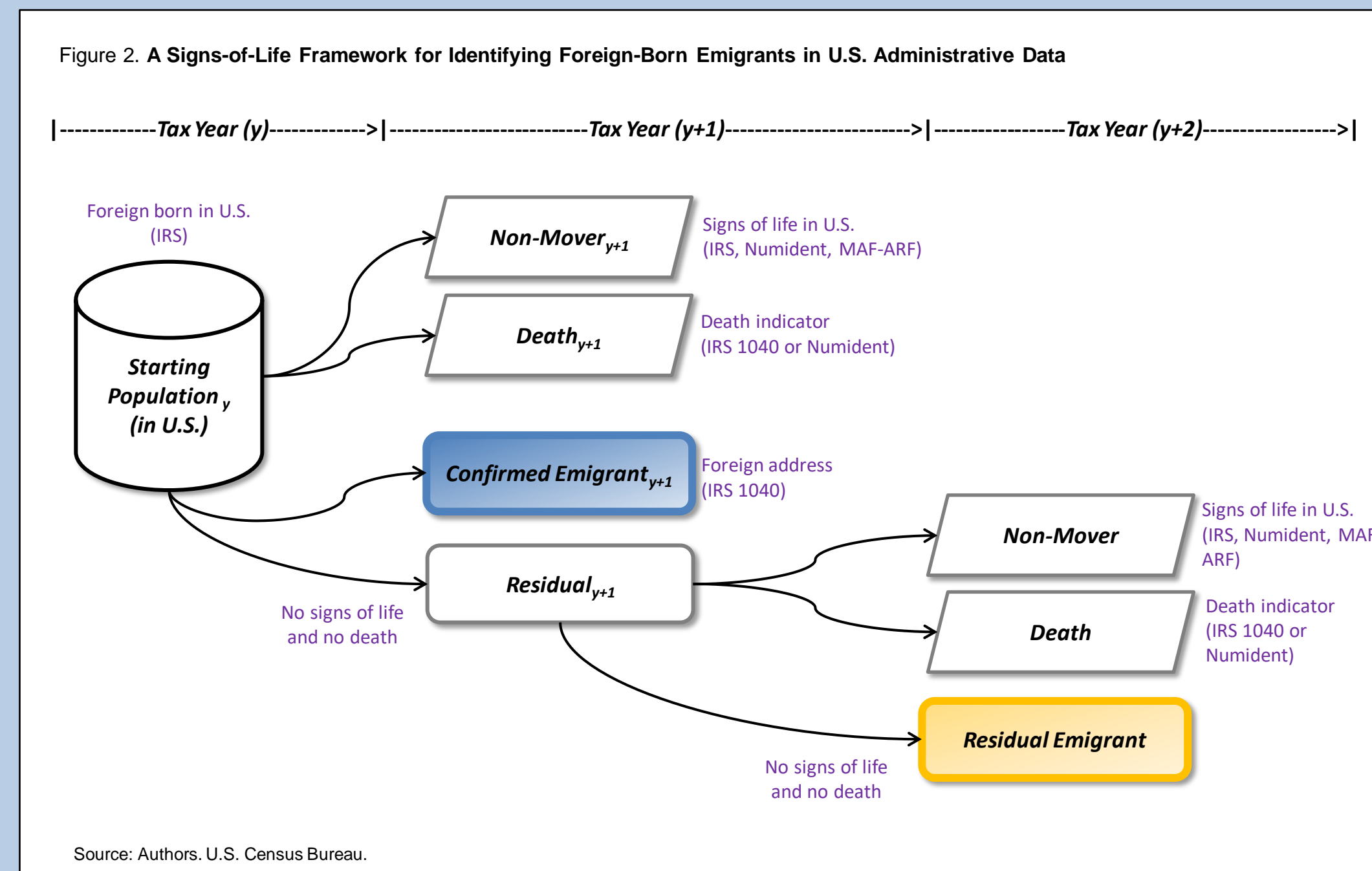
- Geographic variables come from **IRS Form 1040, 1098, 1099, and W2** microdata.
- Demographic variables come from the **Census Numerical Identification file (Numident)**, which is the universe of individuals ever issued a Social Security Number.
- Death indicators come from **IRS Form 1040** and **Numident**.
- Other geographic data come from the **Master Address File-Auxiliary Reference File (MAF-ARF)**.

DATA AND METHODS (CONT'D)

Individual-level linkages across these datasets are made using unique and anonymous Protected Identification Keys (PIKs) assigned to individuals by the U.S. Census Bureau via the Person Identification Validation System (PVS) (Wagner and Layne 2014).

SIGNS-OF-LIFE FRAMEWORK

We first construct a **starting population** of foreign born with Social Security Numbers (SSN) in the U.S. for tax year y . To exclude potential non-residents who were not in the U.S. most of the year, we restrict the population to individuals who appeared in tax data for both the previous ($y-1$) and current (y) tax years.

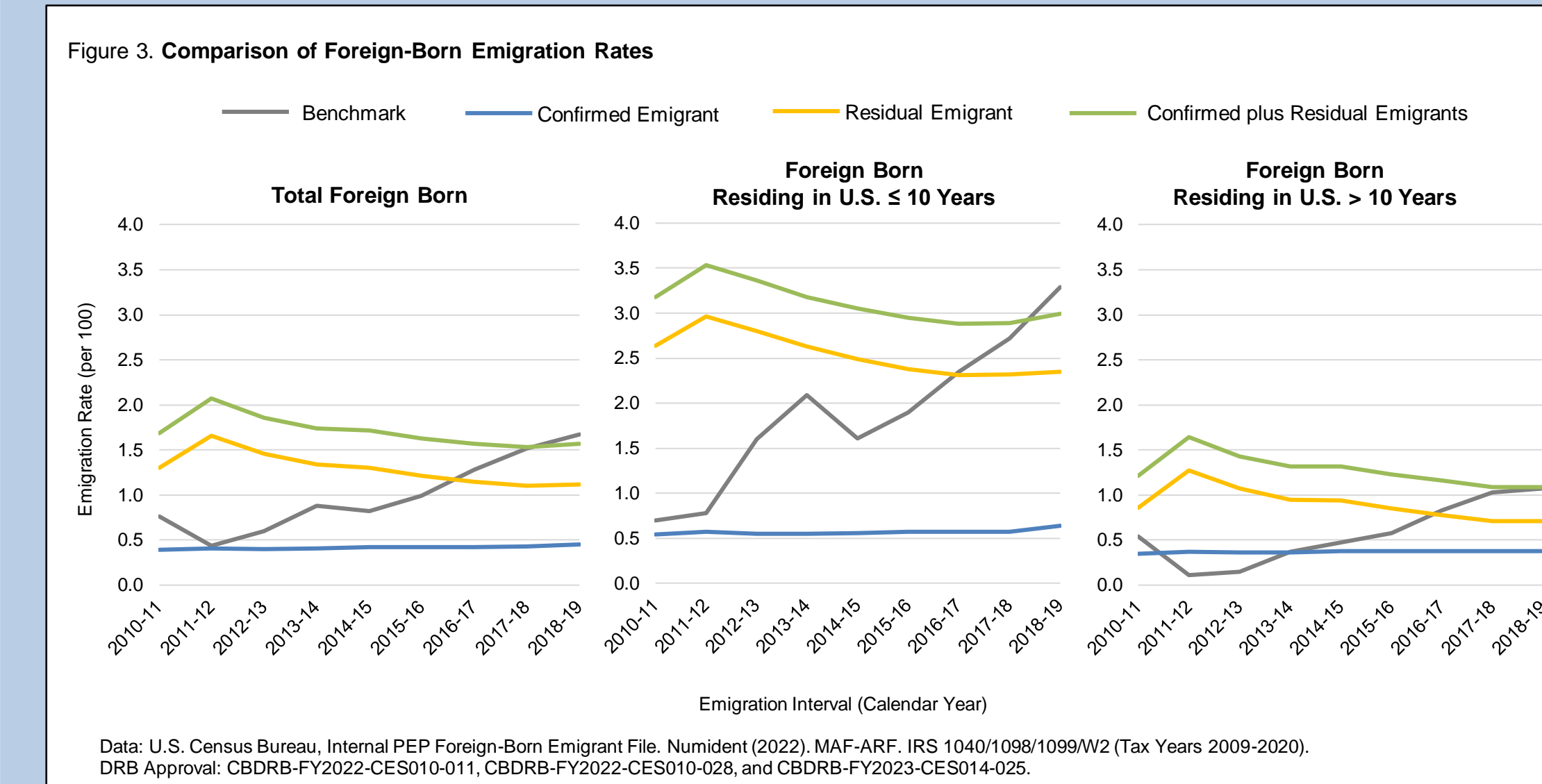


- Non-Movers:** the starting population with U.S. signs of life in the administrative data at $y+1$
- Deaths:** the starting population with an IRS 1040 or Numident death indicator for $y+1$
- Confirmed Emigrants:** the starting population with signs of life at a foreign address (IRS Form 1040 is the only data set in our inventory that records foreign addresses)
- Residual Emigrants:** the starting population with no U.S. signs of life at $y+1$ or $y+2$. The residual in $y+1$ includes both emigrants that had no signs of life for tax year $y+1$ and non-movers who simply had no transactions that year. To distinguish non-movers from emigrants, we allow non-movers an additional year to reappear in the data. If neither death nor U.S. signs of life are detected by $y+2$, we assume the individual left the U.S. and deem them a **residual emigrant**.

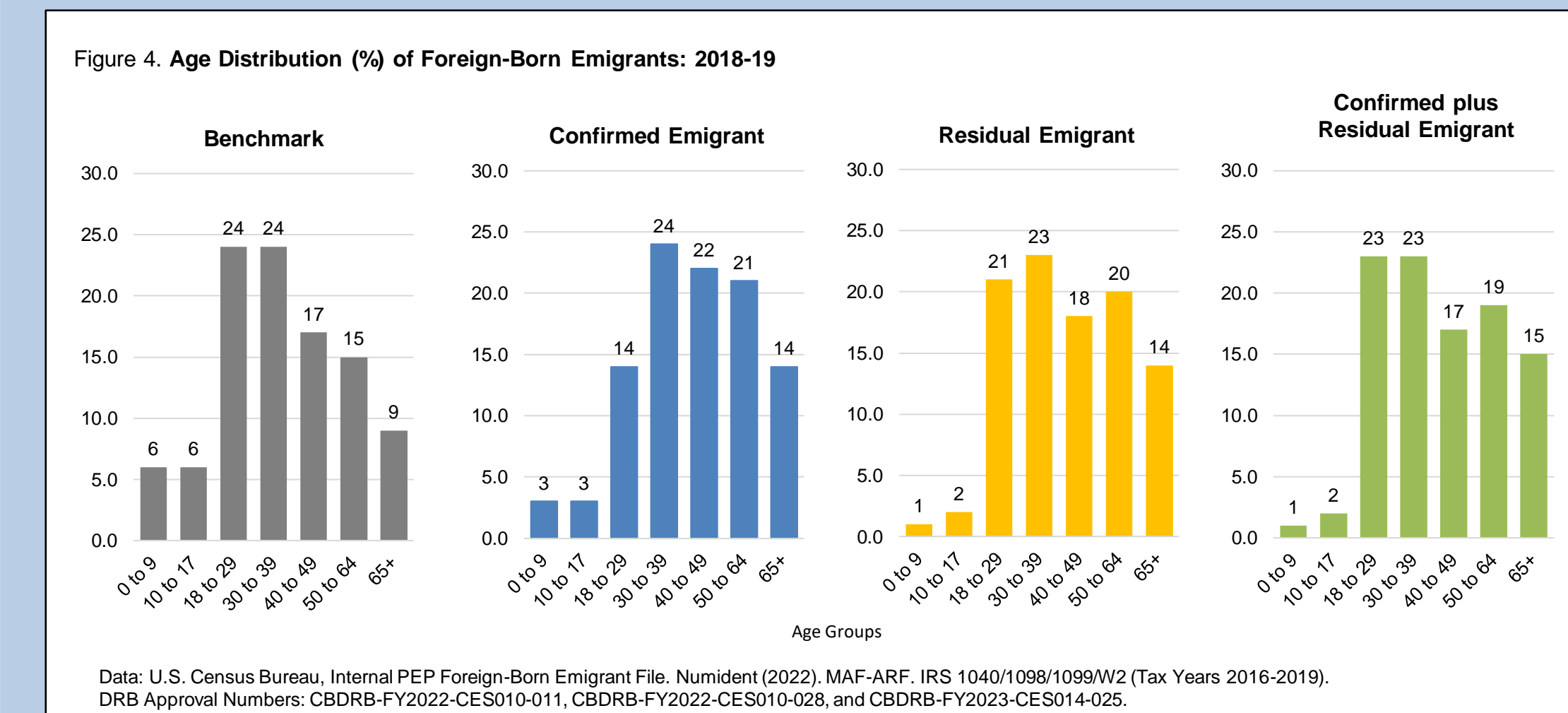
RESULTS

This research analyzes confirmed, residual, and total (confirmed plus residual) emigrants. The **benchmark** represents foreign-born emigration estimates from the PEP non-adjusted¹ ACS-to-ACS Residual Method.²

EMIGRATION RATES

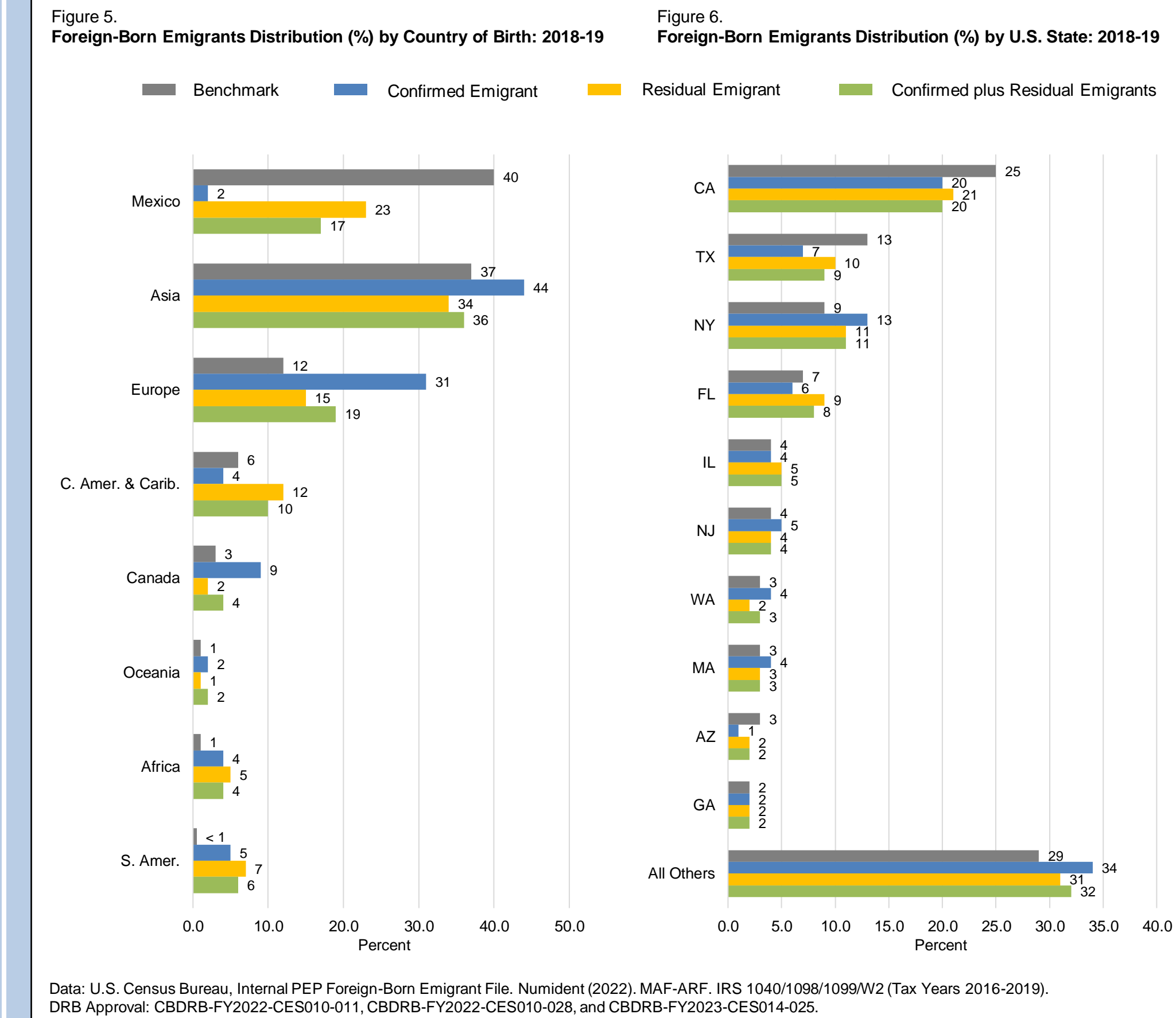


AGE GROUPS



RESULTS (CONT'D)

COUNTRY/REGION OF BIRTH AND STATES



CONCLUSIONS

Administrative data yield less volatile emigration rates than the benchmark and suggest that emigration rates over the previous decade may be relatively flat (Figure 1). However, administrative emigration rates are much higher than rates from previous research, and administrative emigration rate declines contradict benchmark rate increases (Figure 3).

Ongoing research will attempt to understand these rate differences and address **selection bias** in the administrative data.

Measuring the immigrant population in the U.S. using administrative records is challenging. Current universe is limited to subset of foreign born who have SSNs. Mexican born and younger children are underrepresented in the administrative data.

Declining emigration may be real, an artifact of tax policy, or a sign of increasing coverage.