An Analysis of Net International Migration Subnational Distributions Using Administrative Data Compared to the American Community Survey

Megan Benetsky, Esther Miller, Anthony Knapp, Shannon Sabo, Heather King, U.S. Census Bureau

BACKGROUND

The Census Bureau's Population Estimates program releases annual vintage estimates beginning with the most recent decennial census year. One component of these estimates are **Net International Migration (NIM)** flows at the national, state, and county level, by demographic characteristics (age, sex, race, and Hispanic origin). **American Community Survey (ACS)** *proxy universes* are used to distribute the national level total estimate by characteristics. Subnationally, there are year-to-year fluctuations at the county level due to survey sampling and non-sampling error. Also, because we rely on ACS data, inputs can be several years old depending on the geography being estimated. Our proxy universes include immigrants who may have entered the country up to 11 years prior to the survey year. Lagged data and these proxy universes make it difficult to measure recent migration trends (Figure 1).

The **Demographic Characteristics File (DCF)** is produced by the Census Bureau. This file links *current* IRS tax and social security (NUMIDENT) data to decennial census data. The DCF includes geographic and demographic information that could be used to distribute net international migration estimates subnationally.

In this poster, we analyze the feasibility of replacing proxy universes derived from the ACS with distributions from a subset of the DCF, the **Immigrant Demographic Characteristics File (IDCF)**, that includes only foreign-born new entrants. We compare how totals and characteristics are distributed between Vintage 2017, the NUMIDENT, and DCF estimates of immigrants.

V2017 NIM PROXY UNIVERSES

Data:
State: 3-Year Pooled ACS files
County: 5-Year ACS
Proxy Universe Derived from ACS question: *"When did this person come to live in the United States?"*

Universe: Foreign-born immigrants who arrived up to 5 years ago.

Estimate: Apply distributions to national total to estimate totals and characteristics at the state level, and next apply state totals to estimate county totals and characteristics.

Stock: When 3-Year Pooled and 5-Year ACS estimates are used, combined with the proxy universe, subnational estimates and characteristics reflect recent immigrant stock of the foreign-born population rather than annual inflow of immigrants.

Figure 1. Representation of data and proxy universes for foreign-born immigrants at the county level for Vintage 2017

IDCF

Figure 2. IDCF Inputs

NUMIDENT	IRS	Census 2010
• Age	• Geography	• Race
• Sex	• PIK	• Hispanic origin
• Country of birth		• PIK
• Date of entry to file		(Race and Hispanic origin are imputed for the foreign born who entered after 2010.)
• PIK[1]		

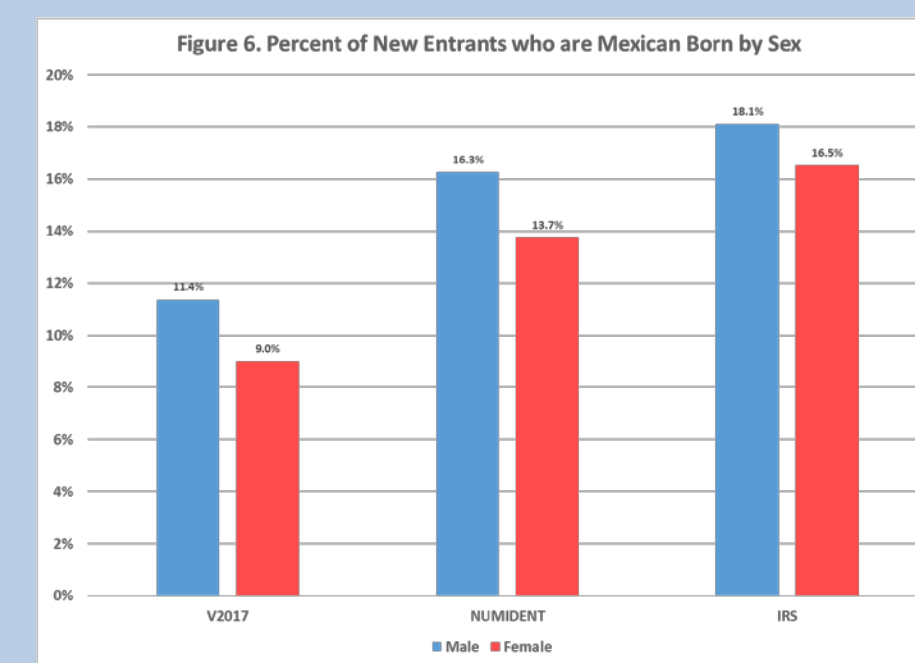[1] Protected Identification Keys (PIKs) that uniquely identify individuals while ensuring confidentiality.
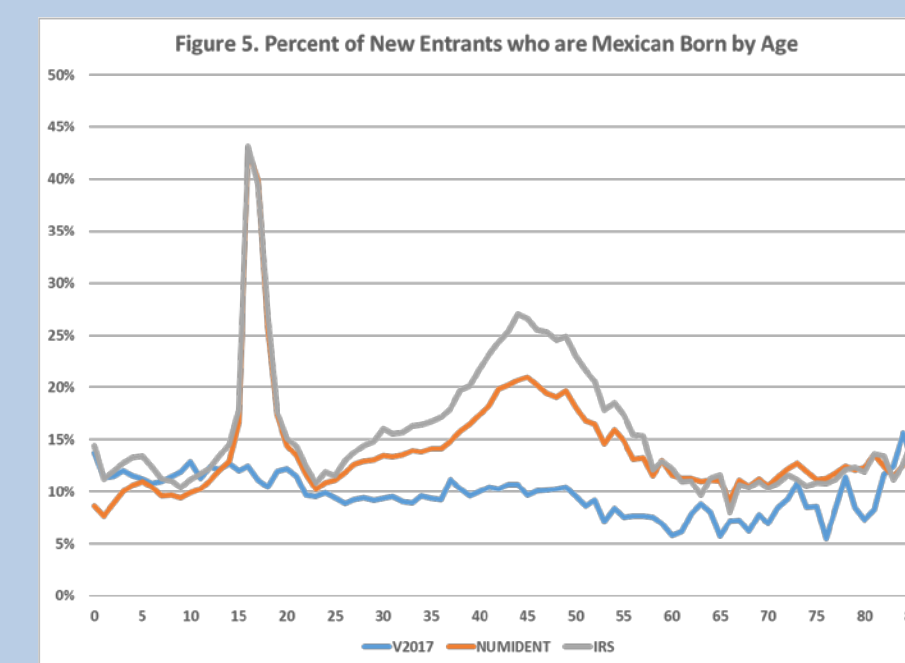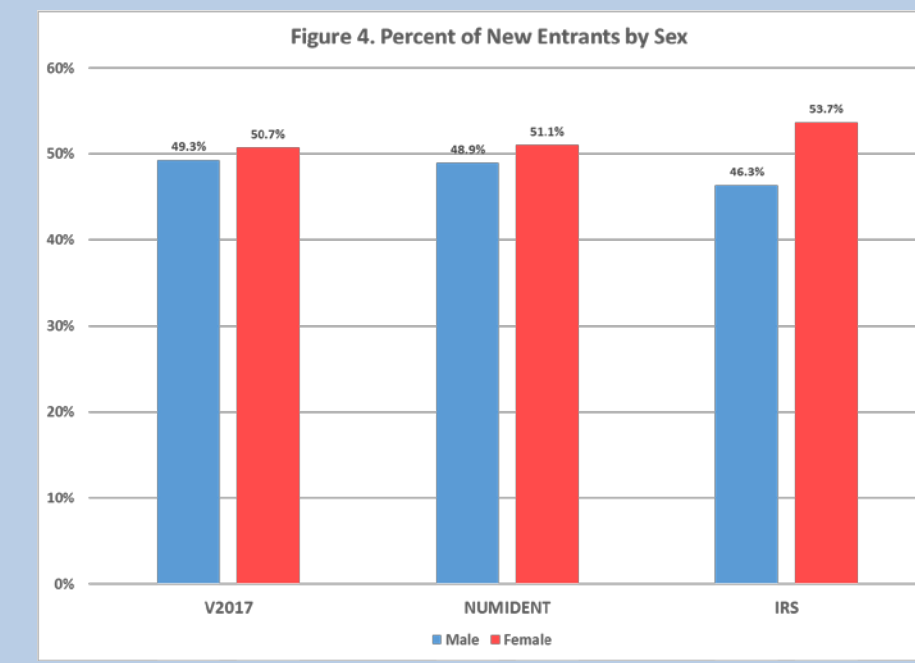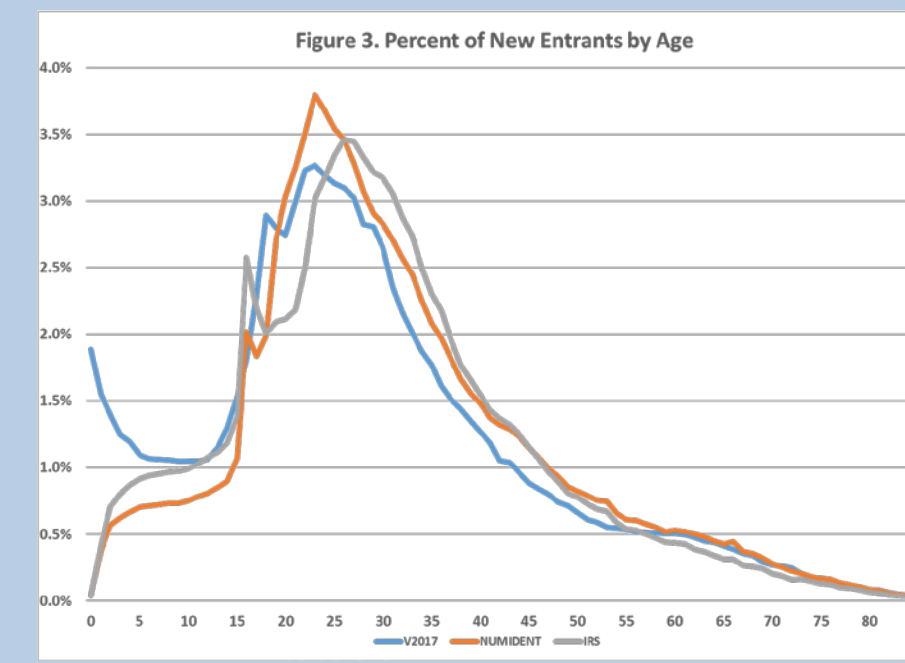
NUMIDENT & IRS UNIVERSES

- The NUMIDENT universe is limited to any person who was assigned a Social Security Number (SSN) between July 1, 2015 and June 30, 2016 (new entrants).
- The IRS universe includes only new entrants who filed taxes in 2016 for tax year 2015. This allows us to use the filer's address from 2016.

We assume that immigrants file for SSNs and taxes in the same year they entered, though we acknowledge this may not be the case for all immigrants, especially recent migrants, irregular migrants, as well as students.

These files are matched by PIK to create the IDCF, which is then compared to NIM data from Vintage 2017, Estimate Year 2017, as the most recent input data for these estimates come from the 2016 ACS, allowing for consistent years of analysis.

NATIONAL ANALYSIS

Figure 3. Percent of New Entrants by Age

Figure 4. Percent of New Entrants by Sex

Figure 5. Percent of New Entrants who are Mexican Born by Age

Figure 6. Percent of New Entrants who are Mexican Born by Sex

Data sources for Figures 3-6:
V2017 – V2017 Estimates, Estimates Year 2017 produced using auxiliary ACS inputs
NUMIDENT – 2016 NUMIDENT that includes foreign-born entrants who entered the U.S. between 7/1/2015 - 6/30/2016
IRS – Tax data filed in 2016

NEW ENTRANTS IN NUMIDENT

Figure 7. Percent of New Entrants who are Mexican Born Appearing in NUMIDENT, Ages 16-18, by Period of Entry
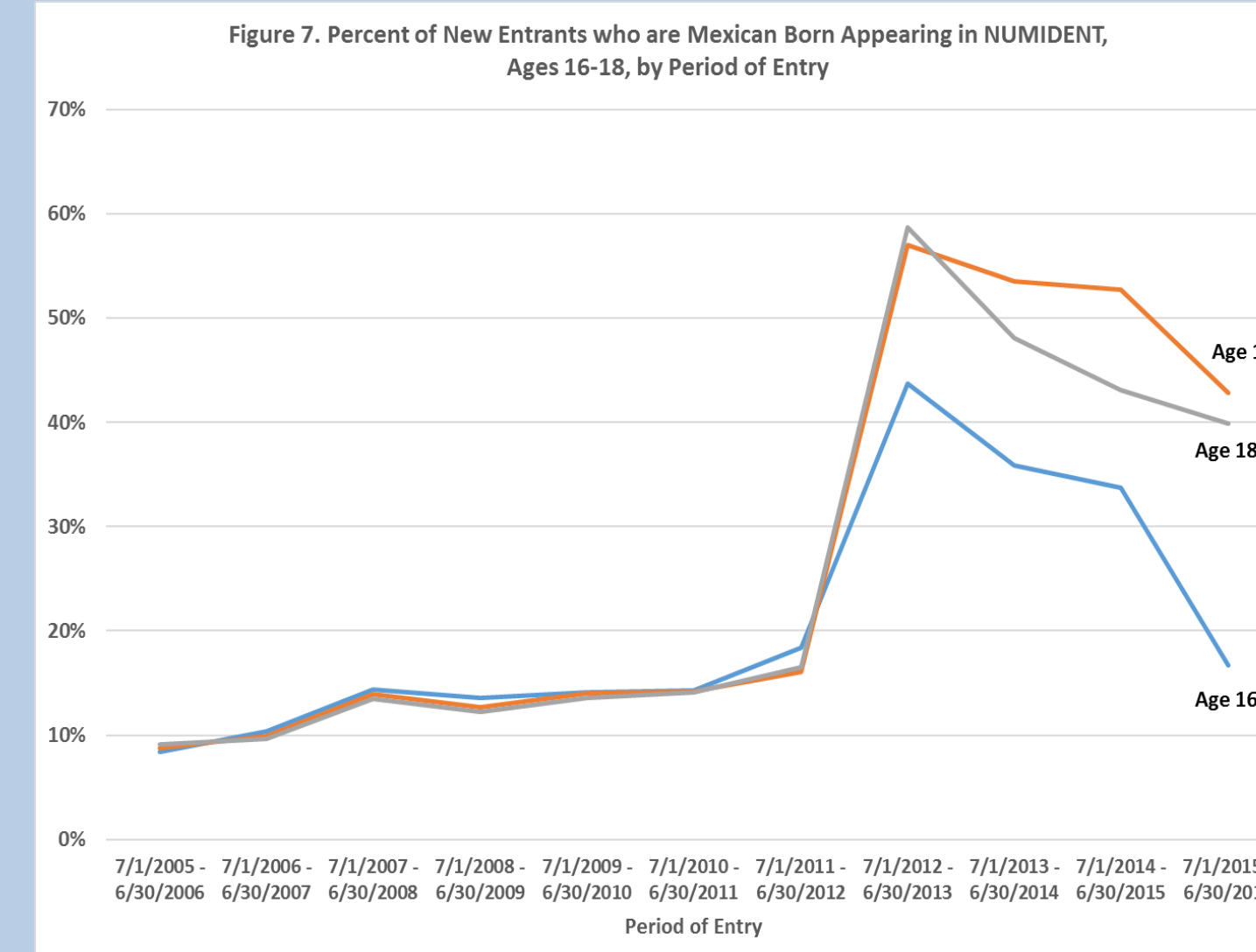
Figure 5 shows a spike in Mexican-born new entrants between the ages of 16 and 18 in the 2016 NUMIDENT and IRS data filed in 2016, while the V2017 data remained relatively stable across age. To look at this pattern for 16-18 year olds historically, we used NUMIDENT files from 2005 through 2016.

Figure 7 shows the percent of new entrants who are Mexican born found in the NUMIDENT by period of entry. The figure indicates that from 2005 to 2011 the trend was mostly constant. The increase in 16-18 year olds began around 2012 and has declined for these ages going forward. In 2016, the percent of newly entered 16 year olds returned to percents observed prior to 2012.
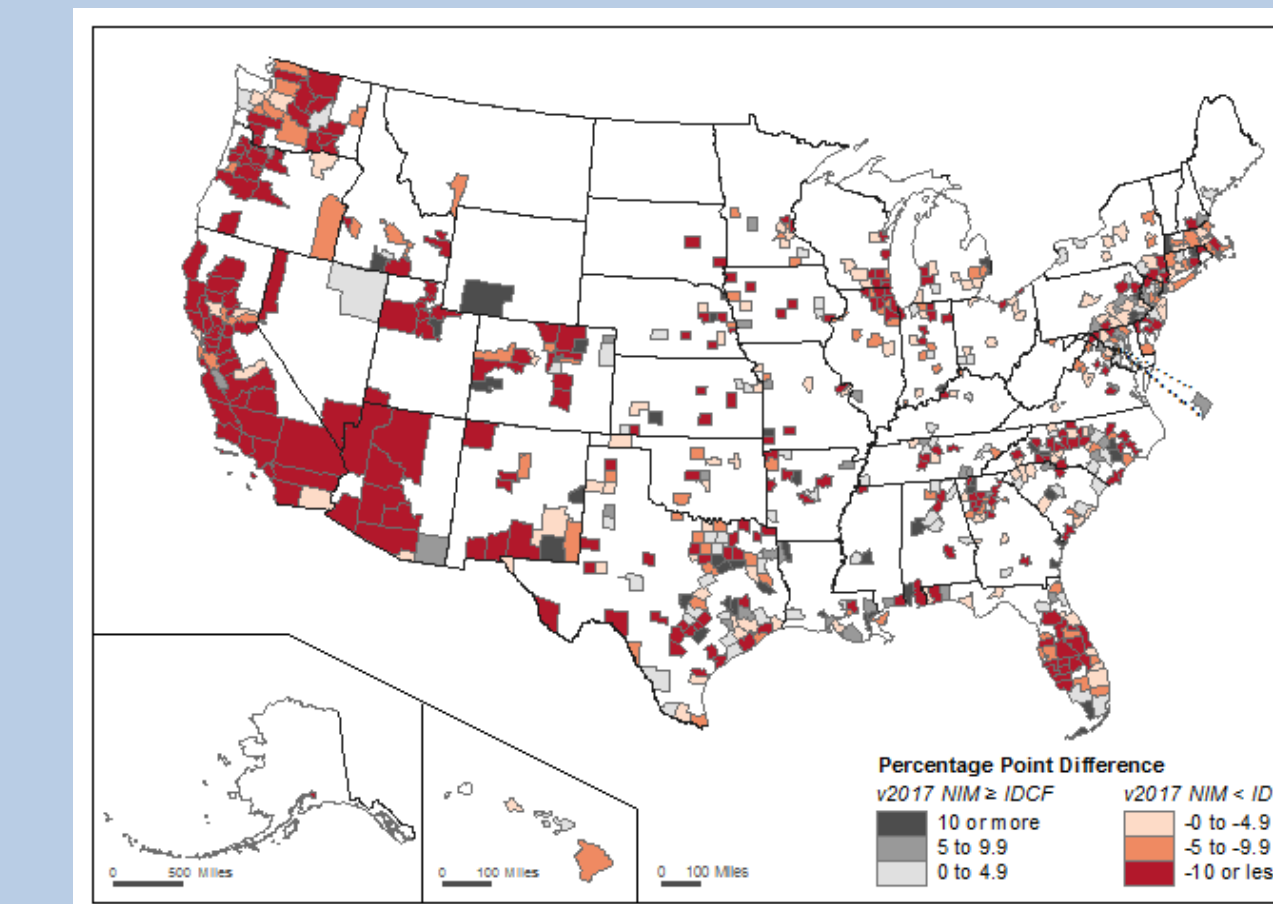
COUNTY ANALYSIS

Figure 8. Percentage Point Difference between NIM and IDCF Data in Asian Alone Immigrant Population by County

Figure 9. Percentage Point Difference between NIM and IDCF Data in Hispanic Immigrant Population by County
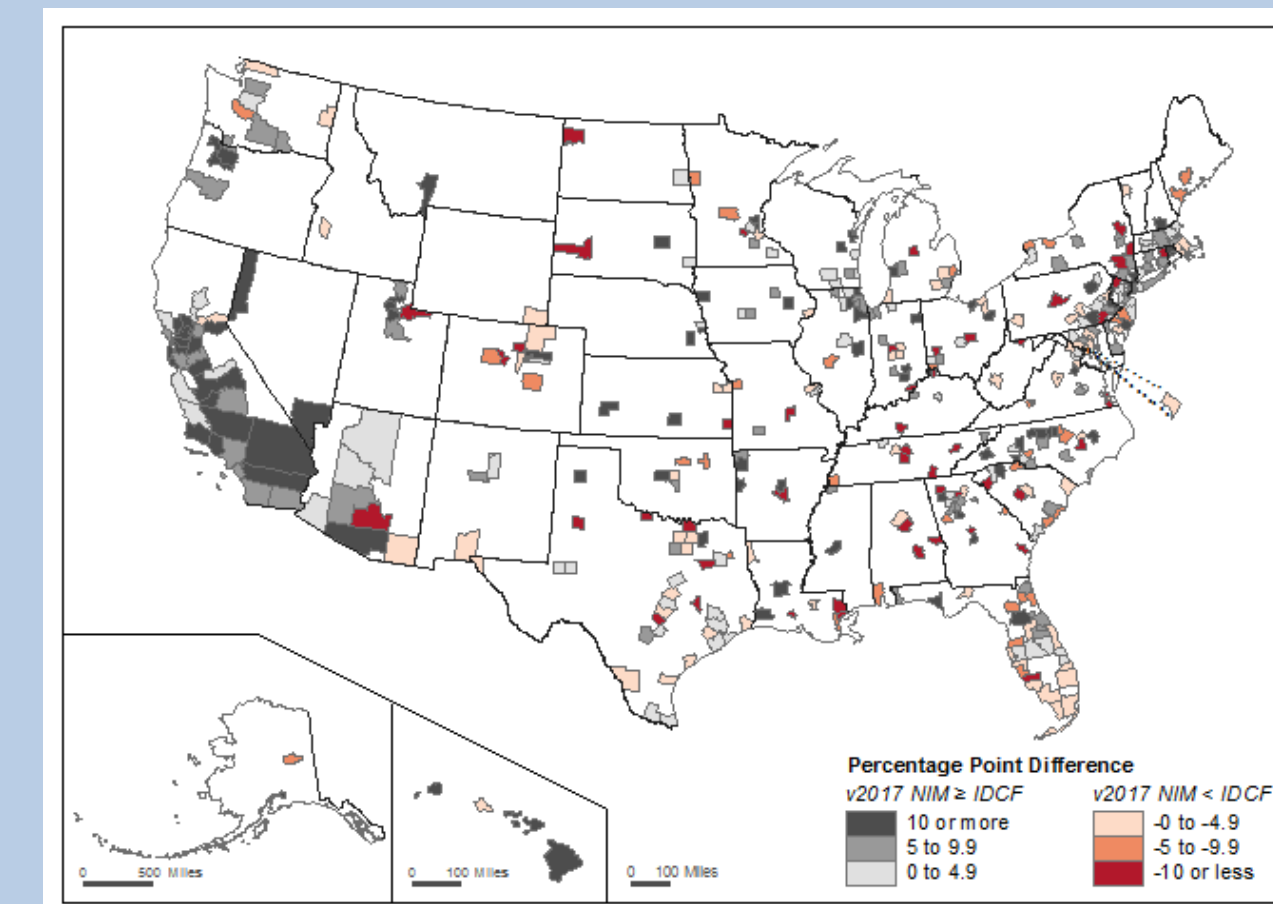
Figures 8 and 9 display the percentage point difference in shares of Asian and Hispanic immigrants by county. First, the percent Asian and Hispanic immigrant population were calculated for each geography using NIM and IDCF data. Percentage point differences in shares of population were calculated to determine differences between datasets.

Sources for Figures 7-10:
V2017 Estimates, Estimates Year 2017 produced using auxiliary ACS inputs
DCF – Matched 2016 NUMIDENT and tax data filed in 2016

Notes:
Figure 9 includes only counties where the Asian immigrant population in the DCF was greater than 20.
Figure 10 includes only counties where the Hispanic immigrant population was at least 10 in NIM data and at least 20 in the DCF.

DISCUSSION

Nationally, the age distribution across V2017, NUMIDENT and IRS data were relatively consistent with one another. The IRS data was 54% female compared to 51% in NIM data. However, there were fewer females when looking only at the Mexican born in V2017, NUMIDENT, and IRS data.

Further analysis of the NUMIDENT shows that the spike in Mexican records in NUMIDENT and IRS data among 16-18 year olds is found from 2012 to 2016 (Figure 7).

Figures 8 and 9 show that V2017 data had larger shares of Asian immigrants while IDCF data had larger shares of Hispanic immigrants. This is especially true in CA counties like Los Angeles, Orange, Riverside, San Bernardino, and San Diego, as well as Maricopa, Pima, and Pinal Counties, AZ. These counties are known to attract both Asian and Hispanic immigrants, but research is needed to understand why V2017 and IDCF data reflect these populations differently.

CONCLUSION

Evaluation of the feasibility of using the IDCF to replace proxy universes used in V2017 has the potential to provide estimates on more recent immigration trends, as well as mitigate some of the issues that arise with survey data. This analysis showed general agreement between V2017 data and the IDCF when looking at age, but for sex, country of origin, and race, more differences emerged.

The definition of new entrants, however, is challenged in the IDCF due to the increase in 16-18 year olds found in the NUMIDENT and IRS data compared to V2017.

In order to determine if the IDCF is a reasonable alternative to the ACS proxy universes to improve data issues that arise from sampling error and possibly recency of the data, the findings in this poster must be further investigated. This includes refining the definition of new entrants in the IDCF and understanding the subnational differences in the Hispanic and Asian populations. These differences may be due to the differences in universes or the definition of new entrants in V2017 and the IDCF, but additional research is needed.