ADEP WORKING PAPER SERIES

Using the Census Bureau's Master Address File for Migration Research

John Sullivan
US Census Bureau

Katie Genadek
US Census Bureau

Using the Census Bureau's Master Address File for Migration Research

John Sullivan, US Census Bureau, john.sullivan@census.gov
Katie Genadek, US Census Bureau, katie.r.genadek@census.gov

Abstract

The U.S. Census Bureau created the geospatial Master Address File (MAF) database as a frame of domiciles with addresses, census geographies, and precise geocodes in the U.S. In addition to the MAF, the Census Bureau links individuals to domiciles in the MAF-Auxiliary Reference File (MAF-ARF) on a yearly basis since 2000. These data are available for research purposes in a restricted-use environment, yet there is limited research using these data and minimal documentation of the files. We address this by describing the MAF and MAF-ARF in detail. We estimate internal migration rates using the MAF-ARF and compare the results to other published estimates on migration, and we perform direct comparisons of location history from migration questions in surveys to MAF-ARF location history at the individual-level. Finally, we use these data to create novel estimates of migration distance for movers in the U.S. Through these analyses, we show the power and potential usefulness of these data for various geospatial research and policy applications.

Migration information in modern person-level data is usually in the form of cross-sectional microdata samples such as the American Community Survey (ACS) or panel surveys with relatively small samples, for example, the Panel Study of Income Dynamics (PSID) or the Current Population Survey (CPS). Aggregate migration flow data is available from the Internal Revenue Service's Statistics of Income (SOI) and covers a large proportion of the U.S. population, but it lacks both geographic and demographic detail (Deward et al. 2022). As migration is a relatively rare event, small microdata samples have left many scholars studying migration unable to explore complexities of recent migration in the United States. Administrative records present an alternate source of migration information with several advantages over traditional data sources such as population level coverage and incorporation in a record linkage infrastructure and with some disadvantages including data collected without a sampling design and not designed for migration measurement (Ernsten et. al. 2018; Foley, Champion and Shuttleworth 2020; Dillon 2021). The Census Bureau makes yearly longitudinal administrative and survey data on residential location for nearly the full population of the U.S. available in a restricted environment for research. While the data have been in development for more than two decades, they are only recently available to researchers and have limited documentation. This paper describes this new and powerful data resource and presents our exploration of migration patterns in the U.S. since 2000.

Background

The U.S. Census Bureau created the Master Address File (MAF) for the 2000 Census. The goal was to create an improved mailing list for the 2000 Census, including non-city style addresses, and integrate the addresses into the Topologically Integrated Geographic Encoding and Referencing system (TIGER®) database. Starting with the list of addresses used for the 1990

Census (the Address Control File), the Census Bureau worked with the U.S. Postal Service

(USPS) and local governments though the Local Update of Census Addresses (LUCA) program,

the MAF became a complete repository for every residential mailing address with location

descriptions used for the 2000 decennial census operations (see 2000 Census procedural history

for full history of the MAF creation (U.S. Census Bureau 2009)). In the two decades that

followed, the MAF has grown to become a rich database of U.S. domiciles with their addresses,

geographic identifiers, time-specific census geographies, and precise geolocations. The MAF

was used for the following two decennial censuses and many other Census Bureau surveys.

      The Geography division within the Decennial Directorate of the Census Bureau manages

and maintains the MAF database. Every year, they create an extract of the database for research

purposes known as the MAF Extract (MAF-X). The MAF-X includes all domiciles in the U.S.

from 2000 to the time of the extract, which are assigned a unique identifier, called the MAFID.

For each MAFID, the MAF-X includes the associated address, physical geographies, census

geographies, and location latitude and longitude. Any survey, census, or administrative

household record that has been assigned a MAFID by the Census Bureau can be linked to the

MAF-X by MAFID to obtain the detailed geographic information for that domicile.

      The Data Acquisition and Curation team at the Census Bureau uses the MAF-X along

with other data to assign yearly locations to individual records, creating the MAF-Auxiliary

Reference File (MAF-ARF). Thus, the MAF-ARF provides an individual-level unique identifier,

or Protected Identification Key (PIK) and a MAFID for all individuals in a given year that are

associated with a domicile in that year in the 50 states and Puerto Rico. The MAF-ARF is

created by finding all PIK and MAFID combinations from seven sources of trusted federal

administrative data[1] where PIKs and MAFIDs have been assigned (Wagner and Layne 2014).

For the years 2000-2011, multiple MAFIDs are retained for each PIK associated with more than

one address in the MAF-ARF. From 2012 to the present, one MAFID is chosen at random when

there are multiple MAFIDs for an associated PIK, so there is only one PIK-MAFID pair per year

in the MAF-ARF.

MAF-ARF Contents and Linkage

Table 1 shows unique record counts for each year of the MAF-ARF from 2000 to 2021.

The MAF-ARF includes hundreds of millions of records in each year corresponding to around

90% of the total population of the United States in any given year. For each year from 2000 –

2021, we link the MAF-ARF to the Census Numident, a person-level extract of the Social

Security Administration's (SSA's') Numerical Identification (Numident) file, to recover

individual dates of death (Finlay and Genadek 2021). Some individuals appear in the MAF-ARF

after a date of death is recorded in the Census Numident, sometimes for multiple years after the

date of death. For example, a PIK could appear in the 2008 MAF-ARF but have a 2007 death

date in the Census Numident because a surviving spouse filed a tax return on behalf of the

decedent. We remove such records and present revised records counts in the MAF-ARF for

individuals alive on January 1st of the reference year.

---

[1] The source files can include IRS 1040 and 1099 files, Selective Service System, the Medicare Enrollment Database, Indian Health Service database, HUD Public and Indian Housing, and HUD Tenant Rental Assistance Certification System (Graham, Kutzbach and Sandler 2017).

Table 1. Master Address File - Auxiliary Reference File Records (2000-2021)

Year	Total US Population	Unique PIKs in the MAF-ARF	Proportion: Unique PIKs/U.S. Population	Unique PIKs Alive in Ref. Year	Proportion: Unique PIKs Alive in Ref. Year/U.S. Population	Mean MAFID per PIK	Median MAFID per PIK	99th percentile MAFID per PIK
2000	281,422,000	251,100,000	0.892	236,800,000	0.841	1.47	1	4
2001	284,969,000	258,000,000	0.905	242,000,000	0.849	1.50	1	4
2002	287,625,000	260,800,000	0.907	243,400,000	0.846	1.47	1	4
2003	290,108,000	266,200,000	0.918	247,100,000	0.852	1.46	1	4
2004	292,805,000	270,300,000	0.923	249,700,000	0.853	1.45	1	4
2005	295,517,000	275,300,000	0.932	253,100,000	0.856	1.46	1	4
2006	298,380,000	260,200,000	0.872	256,300,000	0.859	1.48	1	4
2007	301,231,000	268,400,000	0.891	264,300,000	0.877	1.50	1	4
2008	304,094,000	274,800,000	0.904	270,700,000	0.89	1.50	1	4
2009	306,772,000	276,800,000	0.902	272,700,000	0.889	1.50	1	4
2010	308,746,000	278,700,000	0.903	274,600,000	0.889	1.46	1	4
2011	311,557,000	285,000,000	0.915	280,600,000	0.901	1.48	1	4
2012	313,831,000	294,000,000	0.937	289,400,000	0.922	1	1	1
2013	315,994,000	268,400,000	0.849	266,300,000	0.843	1	1	1
2014	318,301,000	290,100,000	0.911	285,300,000	0.896	1	1	1
2015	320,635,000	294,100,000	0.917	289,200,000	0.902	1	1	1
2016	322,941,000	297,400,000	0.921	292,500,000	0.906	1	1	1
2017	324,986,000	299,800,000	0.923	294,700,000	0.907	1	1	1
2018	326,688,000	303,300,000	0.928	297,100,000	0.909	1	1	1
2019	328,240,000	305,500,000	0.931	300,200,000	0.915	1	1	1
2020	331,449,000	311,900,000	0.941	305,500,000	0.922	1	1	1
2021	332,049,000	312,900,000	0.942	304,600,000	0.917	1	1	1

Notes: Author's calculations from the MAF-ARF. Date of Death (DOD) obtained from Census Numident. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-029 and CBDRB-FY22-ERD002-020. Census 2000 Summary File 1, U.S. Census Bureau, 2001; Census 2011 Summary File 1, U.S. Census Bureau, 2011; Table 1. Intercensal Estimates of the Resident Population by Sex and Age for the United States: April 1, 2000 to July 1, 2010 (US-EST00INT-01) U.S. Census Bureau, Population Division, September 2011. Table 1. Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019 (NST-EST2019-01) U.S. Census Bureau, Population Division, December 2019; Annual Estimates of the

Resident Population for the United States, Regions, States, District of Columbia, and Puerto Rico: April 1, 2020 to July 1, 2023 (NST-EST2023-POP) U.S. Census Bureau, Population Division, December 2023. Table 2. Resident Population for the 50 States, The District of Columbia and Puerto Rico: 2020 Census. U.S. Census Bureau, April 2021.
Total US Population estimates do not include residents of Puerto Rico or Island Areas.

From 2000 to 2011, the MAF-ARF includes all PIK-MAFIDs pairs that were found in using the administrative sources, thus an individual (PIK) could appear in an annual MAF-ARF file at multiple locations (MAFIDs). The records are not dated and when a PIK is associated with multiple MAFIDs, the various residences cannot be sorted in time. As can be seen in Table 1, in each year from 2000 to 2011, the median number of MAFIDs per PIK is 1 and the 99th percentile is 4. The mean is usually around 1.4, demonstrating that while the majority of PIKs are associated with a single MAFID, some are associated with multiple MAFIDs.

The MAF-ARF is created using administrative records, not a sampling frame. Thus, the inclusion probabilities are unknown and the sample of individuals within the MAR-ARF is not random for two primary reasons. First, the linkage process matches federal records to a database of SSA's Social Security Numbers (SSNs) and Internal Revenue Service's (IRS's) Income Tax Identification Numbers (ITINs), as such, the universe is limited to SSN and ITIN holders. The second reason is that the federal administrative records used to produce the MAF-ARF do not cover the entire population. It is not certain that a person living in the U.S. will appear in at least one of the federal administrative records source files.

To further understand which individuals are included and excluded from the file, we analyze the age and race distributions of the individuals in the MAF-ARF and compare it to the decennial census demographic breakdowns in 2000, 2010, and 2020. Age in each year is determined from date of birth in the Census Numident and is available for nearly all individuals in the MAF-ARF. Race information is available for a more limited set of MAF-ARF records and is measured using reports from the 2000 or 2010 Census and the Census Numident. These results are presented in Tables 2 and 3, respectively. The MAF-ARF is fairly representative of the US population and includes very large samples across the age distribution and for all racial groups.

However, the MAF-ARF underrepresents children, particularly in 2000, where 4.2% of the

MAF-ARF sample are children under age 5, compared to 6.8% in the 2000 Census. Differences

between the MAF-ARF and census shares of children under the age of 15 decrease over time.

The MAF-ARF and censuses have very similar proportions of records in ages 25-79, but the

MAF-ARF overrepresents adults 85 years and older.

Table 2. Age Distribution in MAF-ARF Sample, Decennial Census and ACS (2000-2020)

Age Group (Proportion)	2000		2010		2020	
	MAF-ARF	Decennial	MAF-ARF	Decennial	MAF-ARF	ACS
1-4 years	0.042	0.068	0.045	0.065	0.043	0.06
5-9 years	0.057	0.073	0.055	0.066	0.057	0.06
10-14 years	0.058	0.073	0.056	0.067	0.061	0.065
15-19 years	0.068	0.072	0.068	0.071	0.064	0.065
20-24 years	0.086	0.067	0.093	0.07	0.069	0.067
25-29 years	0.079	0.069	0.083	0.068	0.071	0.071
30-34 years	0.077	0.073	0.07	0.065	0.07	0.068
35-39 years	0.082	0.081	0.068	0.065	0.066	0.065
40-44 years	0.078	0.08	0.068	0.068	0.061	0.061
45-49 years	0.068	0.07	0.073	0.074	0.061	0.063
50-54 years	0.058	0.063	0.069	0.072	0.063	0.064
55-59 years	0.045	0.048	0.06	0.064	0.067	0.067
60-64 years	0.036	0.038	0.052	0.055	0.064	0.062
65-69 years	0.034	0.034	0.04	0.04	0.056	0.053
70-74 years	0.035	0.032	0.03	0.03	0.046	0.041
75-79 years	0.033	0.026	0.025	0.024	0.032	0.028
80-84 years	0.026	0.018	0.021	0.019	0.022	0.019
85+ years	0.04	0.015	0.026	0.018	0.028	0.02
N	305600000	281421906	340700000	308745538	308800000	326569308

Notes: Author's calculations from the MAF-ARF. Date of Death (DOD) obtained from Census Numident. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY22-ERD002-010 and CBDRB-FY22-ERD002-020. Census 2000 Summary File 1, U.S. Census Bureau, 2001; Census 2011 Summary File 1, U.S. Census Bureau, 2011; Table DP05 ACS Demographic and Housing Estimates, 2016-2020 5-Year Estimates, U.S. Census Bureau.

The MAF-ARF is generally racially representative of the U.S. population (Table 3). From

2000 to 2010, the MAF-ARF slightly underrepresents non-Hispanic whites relative to the

Censuses (65% and 69%, respectively) and slightly overrepresents Hispanics of any race (15% and 13%, respectively). Blacks, Asians and AIAN/PI are similarly represented in the MAF-ARF and Censuses.

Table 3. Racial Distribution in MAF-ARF Sample, Decennial Census (2000-2010)

Racial Distribution (Proportion)	2000		2010	
	MAF-ARF	Decennial	MAF-ARF	Decennial
White NH	0.65	0.69	0.62	0.64
Black/Af. Am. NH	0.12	0.12	0.13	0.13
Other/Multi-racial NH	0.022	0.014	0.038	0.014
Asian NH	0.041	0.038	0.045	0.049
Hispanic Any Race	0.15	0.13	0.17	0.16
Am. Indian/PI	0.008	0.01	0.008	0.01
N	245800000	281422000	278500000	308746000

Notes: Author's calculations from the MAF-ARF. Date of Death (DOD) obtained from Census Numident. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY22-ERD002-020. Census 2000 Summary File 1, U.S. Census Bureau, 2001; Census 2011 Summary File 1, U.S. Census Bureau, 2011.

Measuring Migration in the U.S.

Though the MAF-ARF is not designed to collect direct reports of a residential move from a representative sample of respondents, it offers opportunities for novel estimates of geographic mobility in the United States. To estimate geographic mobility, we use the PIK to link the same individuals across years of the MAF-ARF. We then compare the MAFID location of individuals in adjacent years and consider a change in locations in the MAF-ARF to constitute a residential move. This strategy will capture any changes in residential location, from the shortest distance moves (for example, from one apartment unit to another within the same building) to moves between states, however, we cannot observe when an individual moves to another country. From

2000-2011, when a PIK can be associated with more than one MAFID, we randomly choose one among multiple PIK-MAFID pairs to be consistent with the data available after 2011.

In Table 4, we present year over year proportions of MAF-ARF records for which we observe a change in MAFIDs (representing a residential move). The first estimate is unadjusted and based on the full sample of MAF-ARF records with non-missing location information at time t and $t+1$. The second estimate is based on a sample limited to MAFIDs with a valid "unit status" code in the MAF and removes any MAFIDs associated with a PO BOX or with incomplete address information. Though the MAF is carefully and consistently maintained by the Census Bureau's Geography division, occasionally, the same residential location will appear under multiple MAFIDs. Further, we adjust this estimate to remove very short distance moves (less than 500ft)[2]. Very short distance moves do occur, as individuals may change units in the same apartment complex or move a few houses away on the same street. However, other short distance moves measured in the MAF-ARF are likely due to the same housing unit appearing in the MAF multiple times under different MAFIDs. These two estimates provide an upper and lower bound, assuming all short distance moves are valid and assuming all short distance moves are invalid, respectively. In the unadjusted estimate, between 21.5%-27.9% of PIKs in the MAF-ARF appear at a different MAFID then they did in the prior year, with an average of 24.1% moving each year over the entire period. The adjusted estimate is generally lower than the unadjusted estimate, by between 10% and 19%.

---

[2] From the original estimates, we find the distance of moves for housing units with complete latitude and longitude information (a subset of all units in the MAF). We use the proportion of moves less than 1/20th of a mile to adjust the original count of moves. For example, if 6% of moves in a given year are less than 1/20th of a mile, we adjust the original move count down by 6% and then recalculate the mobility rate using the denominator for all valid records, regardless of whether their associated MAFIDs have complete latitude and longitude. This assumes that the proportion of short distance moves is the same in the sample of units with and without complete latitude and longitude information.

Table 4. Geographic Mobility in MAF-ARF (2000-2021)

Year	Unadjusted Estimate			Adjusted Estimate		
	Potential mover population	Movers	Proportion Moved	Potential mover population	Movers	Proportion Moved
2000 - 2001	224,600,000	53,970,000	0.240	210,600,000	42,890,000	0.204
2001 - 2002	228,700,000	54,360,000	0.238	216,300,000	44,440,000	0.206
2002 - 2003	231,100,000	54,010,000	0.234	219,200,000	44,410,000	0.203
2003 - 2004	234,600,000	53,930,000	0.230	223,500,000	44,950,000	0.201
2004 - 2005	237,500,000	55,440,000	0.233	226,800,000	46,340,000	0.204
2005 - 2006	240,700,000	56,660,000	0.235	230,500,000	47,900,000	0.208
2006 - 2007	244,900,000	61,000,000	0.249	232,700,000	48,920,000	0.210
2007 - 2008	254,000,000	60,570,000	0.239	241,400,000	49,800,000	0.206
2008 - 2009	258,700,000	61,830,000	0.239	246,700,000	50,240,000	0.204
2009 - 2010	260,900,000	59,750,000	0.229	250,800,000	50,150,000	0.200
2010 - 2011	260,100,000	66,920,000	0.257	248,600,000	51,860,000	0.209
2011 - 2012	270,400,000	71,850,000	0.266	256,200,000	58,860,000	0.230
2012 - 2013	254,800,000	71,130,000	0.279	258,300,000	60,930,000	0.236
2013 - 2014	254,400,000	54,760,000	0.215	260,800,000	56,430,000	0.216
2014 - 2015	274,400,000	60,460,000	0.220	266,600,000	51,750,000	0.194
2015 - 2016	278,200,000	66,300,000	0.238	270,300,000	56,690,000	0.210
2016 - 2017	281,300,000	70,440,000	0.250	273,400,000	61,320,000	0.224
2017 - 2018	283,500,000	67,560,000	0.238	275,800,000	58,120,000	0.211
2018 - 2019	285,700,000	69,780,000	0.244	278,300,000	59,730,000	0.215
2019 - 2020	289,000,000	70,700,000	0.245	280,800,000	59,640,000	0.212
2020 - 2021	291,400,000	70,300,000	0.241	280,500,000	57,690,000	0.206

Notes: Author's calculations from the MAF-ARF. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-029 and CBDRB-FY23-ADEP001-008.

This is a higher proportion of movers than recorded in other government sources of geographic mobility information, which generally find that between 10% and 15% of Americans move each year between 2000 and 2019 (Molloy and Smith 2019). Figure 1a compares estimates of 1-year geographic mobility made from the MAF-ARF to estimates from the CPS. For 2000-2011, when a PIK may be listed at more than one MAFID, we present five mobility estimates. Table 5 presents the definitions of these estimates. For the "Select One" estimate, we randomly choose among multiple locations, creating a comparable measure to the 2012-2021 period. In the "Any Match" estimate, when we observe at least one identical residential location between t and t + 1, regardless of how many other addresses may be present, we consider this as a stay (the most conservative estimate). Conversely, the "Any Non-Match" estimate (least conservative estimate) treats any non-match in location as a move (even if a matching location is present among multiple locations). For the "Single Records" estimate, we drop all records with more than a single MAFID. The "Single Records" and "Any Match" estimates both trend near the CPS estimate. However, the "Any Non-Match" and "Select One" estimates trend considerably higher than the CPS. The "Select One" estimate exhibits more year-to-year variability than the CPS and does not exhibit the slow decline in geographic mobility, but rather, remains relatively stable over the period 2000 to 2021.

Figure 1b plots estimates for the "Select One" and "Any Match" measures. As expected, the adjusted measures are lower than the unadjusted measures. The adjusted "Select One" estimates remain higher than the CPS estimates, even when assuming all very short distance moves are data artefacts. The adjusted "Any Match" estimates closely approximate the CPS estimates, though they begin to diverge from CPS starting in 2010.

Table 5. Migration Estimate Definitions

Estimate	Definition	Years
CPS	Reference estimate	2000-2021
Any Non-match	Any observed location difference counted as a move	2000-2011
Single Records	Sample limited to PIKs associated with only one MAFID	2000-2011
Any Match	Any observed location continuity counted as a stay	2000-2011
Any Match (adjusted)	Any Match criteria, MAF valid sample and no short distance moves	2000-2011
Any Match (not adjusted)	Any Match criteria, MAF valid sample and no distance restrictions	2000-2011
Select One	One MAFID chosen randomly when multiple are present	2000-2021
Select One (adjusted)	Select One criteria, MAF valid sample and no short distance moves	2000-2021
Select One (not adjusted)	Select One criteria, MAF valid sample and no distance restrictions	2000-2021

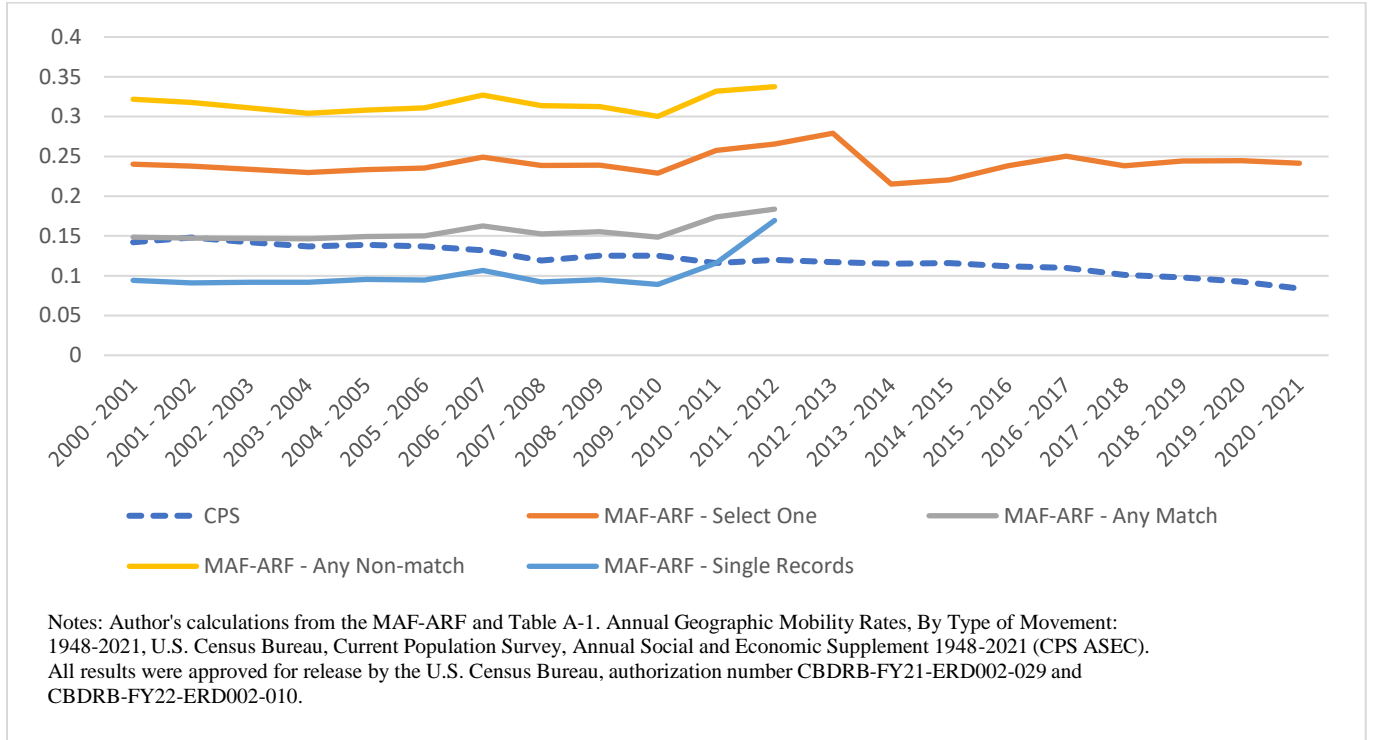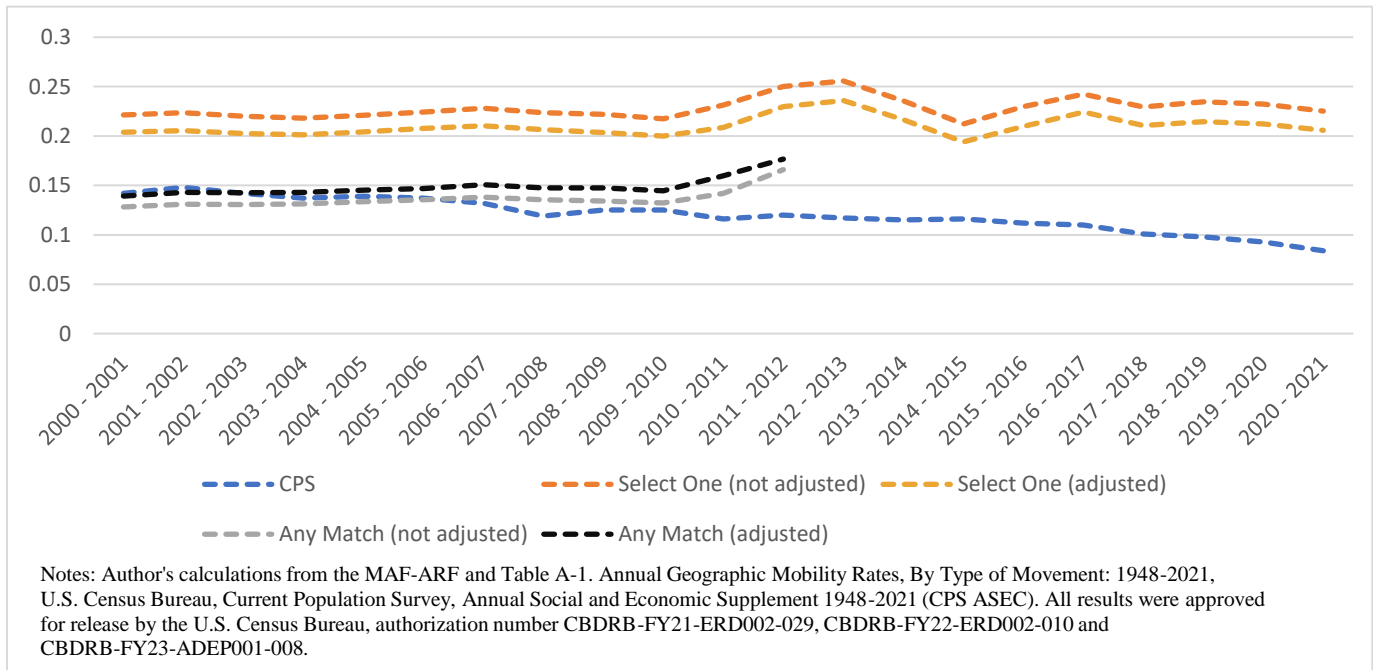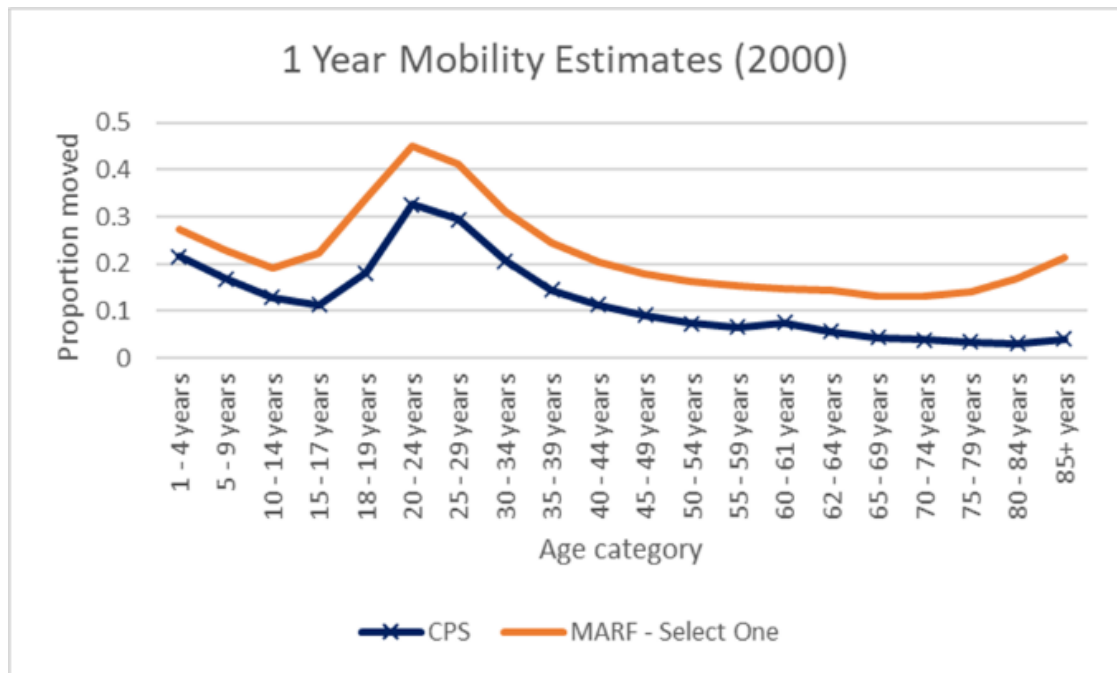Figure 1a. Geographic mobility estimates from MAF-ARF and CPS (2000-2021)

Notes: Author's calculations from the MAF-ARF and Table A-1. Annual Geographic Mobility Rates, By Type of Movement: 1948-2021, U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement 1948-2021 (CPS ASEC). All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-029 and CBDRB-FY22-ERD002-010.

Figure 1b. Adjusted geographic mobility estimates from MAF-ARF (2000-2021)

Notes: Author's calculations from the MAF-ARF and Table A-1. Annual Geographic Mobility Rates, By Type of Movement: 1948-2021, U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement 1948-2021 (CPS ASEC). All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-029, CBDRB-FY22-ERD002-010 and CBDRB-FY23-ADEP001-008.

Mobility estimates from the CPS find that around 15% of Americans moved in 2000, with this figure declining to just below 10% in 2021. While the MAF-ARF estimates for the entire period (using the Select One estimate) are higher than the CPS estimates, the migration patterns by age and race are very similar across sources. The familiar age gradient of residential mobility (Johnson et. al. 2005) is clearly present in both sources and shown in Figures 2a and 2b for 2000 and 2019, respectively.
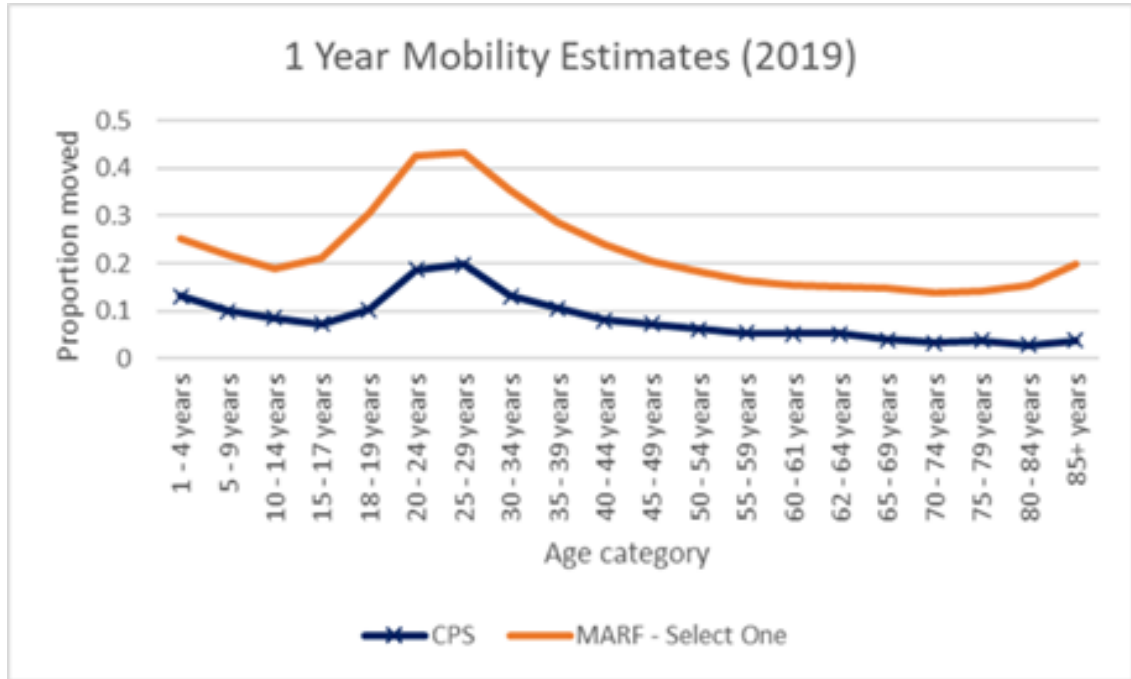
Figure 2a. 1-Year Mobility Estimates (MAF-ARF and CPS 2000)

Notes: Author's calculations from the MAF-ARF and Table A-1. Annual Geographic Mobility Rates, By Type of Movement: 19482021, U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement 19482021 (CPS ASEC). All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-029.

Figure 2b. 1-Year Mobility Estimates (MAF-ARF and CPS 2019)

Notes: Author's calculations from the MAF-ARF and Table A-1. Annual Geographic Mobility Rates, By Type of Movement: 19482021, U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement 19482021 (CPS ASEC). All results were approved for re lease by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-029.

Estimates of geographic mobility from the MAF-ARF are also larger than those from other sources at higher geographic levels. For example, estimates of moves from the CPS and SOI data find that around 3%-6% of Americans move between counties in any year and 1.5% to 3% move between states (Molloy and Smith 2019). In the MAF-ARF, we find slightly higher mobility for between-county moves, with around 8% of the sample changing counties, but similarly to the CPS and SOI estimates, find that around 3%-4% of the MAF-ARF records move between states in a given year between 2000 and 2021.

Individual-Level Comparison

To assess the quality of location information in the MAF-ARF, we link individuals in the MAF-ARF at the person-level to each year of the 2005-2019 ACS and compare their residential

locations. For years when the MAF-ARF allowed multiple MAFIDs per PIK, the MAFID we

selected may have differed from that in which a respondent received an ACS questionnaire.

Panel A in Figure 3 shows that there is a high degree of location agreement between the ACS

and the MAF-ARF; between 75% and 80% of records appear at the same housing unit (MAFID)

in the MAF-ARF and ACS. Nearly, 95% appear in the same county. Discrepancies between

locations in the MAF-ARF and ACS could be attributed to error and measurement inconsistency.

Further, different individuals may have been assigned the same PIK in the ACS and MAF-ARF

data. Location disagreement may also arise from legitimate changes in address occurring

between the time the ACS questionnaire was completed, and the data recorded in the MAF-ARF

was produced.

We also use reported migration information from the ACS as a check on moves observed

in the MAF-ARF. The ACS asks respondents whether they lived in the same place or a different

place one year ago. We compare the MAFID changes we observe in the MAF-ARF to reported

moves in the ACS to measure the extent to which moves are reported in the ACS. We make this

comparison two ways, first by calculating the proportion of ACS respondents for whom we

observe a move in the MAF-ARF that also reported a move in the ACS, and second, by

calculating the proportion of reported ACS moves that are also observed in the MAF-ARF. Panel

B in Figure 3 shows these two proportions for all ACS records matched to the MAF-ARF and

then, in order to limit measurement issues related to timing, for a sample limited to ACS

interviews that occurred in January. The dots toward the left of the figure show that only about

one third of moves in the MAF-ARF have a corresponding reported move in the ACS. This

improves slightly when we look at only January interviews. On the right, we find that about two

thirds of moves reported in the ACS have a corresponding move in the MAF-ARF, again this

improves slightly when we consider only January interviews. Our results are generally like those reported by Foster, Ellis and Fiorio (2019), who corroborate 56.8% of moves in the ACS 2010-2013 with IRS records.

Figure 3a
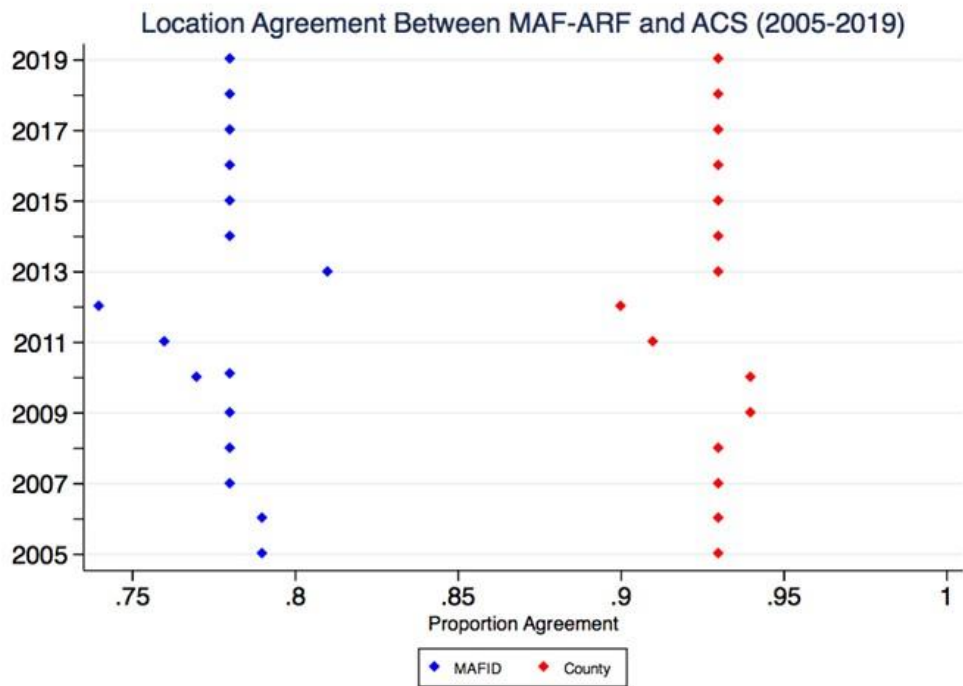
Notes: Author's calculations from the MAF-ARF, 2010 Census and ACS. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY22-ERD002-007.

Comparing moves in the MAF-ARF and ACS (2005-2018)
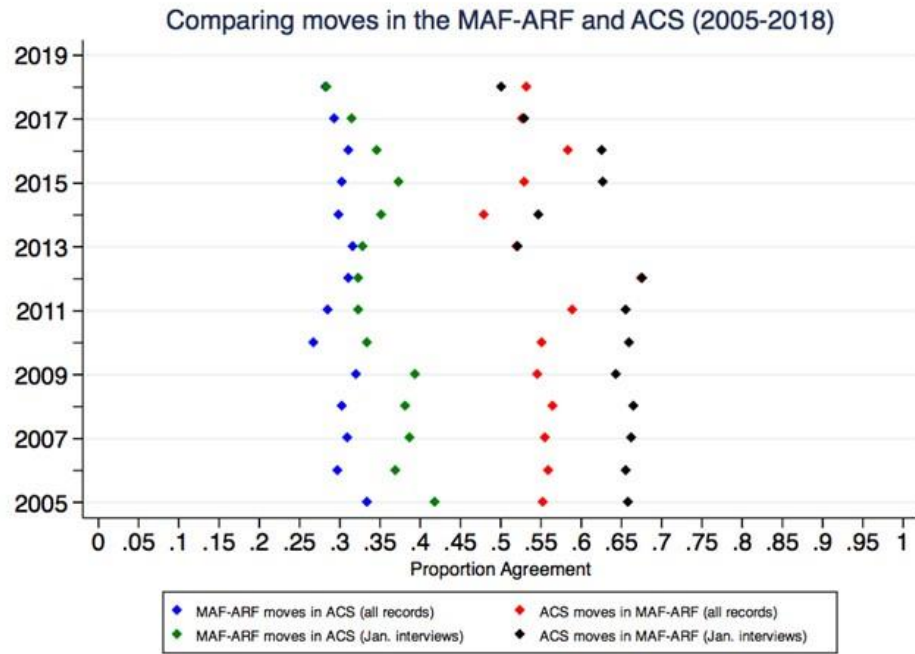
Notes: Author's calculations from the MAF-ARF and ACS. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY22-ERD002-007.

Distance of Residential Moves

Due to lack of detailed geography in most sources of mobility data, studies of precise move distance are uncommon. However, prior work using administrative records sources (SOI) and survey sources (ACS) shows that local or short distance moves (often defined as moves within county or census tract) are far more common than moves between counties, states, or regions. Estimates of interstate moves tend to be highest (2.5%-3%) in the SOI data, lowest in the CPS (1.5%-3%) and between the SOI and CPS in the ACS (Molloy and Smith 2019). Moves between counties are more common with around 3%-5% of the US population changing counties in any year over the period 2000-2019 in the CPS and 5%-6% of people doing so in the SOI data (Molloy and Smith 2019). Further, around 65% of moves were within the same county in 2019, 17% between counties in the same state and only 14% between states (Frost 2020).

We utilize the detailed geographic information available in the MAF-X to calculate move distance for movers in each year. For individuals with different MAFIDs in two consecutive years, we calculate the Euclidean distance of a residential move as the distance between the latitude and longitude coordinates of the two MAFIDs. Table 6 shows the distribution of move distance from 2000 to 2021 for the sample of Select One moves used in the adjusted estimates in Table 4 (limited to moves longer than 500 feet). Most moves are over a relatively short distance, with the modal move being 6-7 miles for the period 2000 to 2021. The distribution is right-skewed, as longer distance moves pull the average move to between 145 and 165 miles. Move distance in miles provides a unique perspective on how Americans move. For example, consider that 10% of the moves we observe are between 460 and 575 miles. While interstate moves cross meaningful administrative borders, they may vary in distance depending on proximity to a state border. Moves of hundreds of miles, even if within the same state, are costly and may more accurately describe moves to different communities than interstate moves alone. Table 7 shows the distribution of move distance for the sample used in the unadjusted estimates (very short distance moves not removed).

Table 6. Distance in Miles of Moves greater than 500 ft in MAF-ARF (2000-2021)

Year	N	Mean	SD	10th Pct.	Median	90th Pct.
2000 - 2001	34320000	154.7	421.9	0.63	6.4	508.8
2001 - 2002	35430000	150.9	416.9	0.61	6.4	484.6
2002 - 2003	35530000	147.5	412.3	0.62	6.3	464.4
2003 - 2004	36260000	146.8	411.3	0.64	6.4	460.1
2004 - 2005	37530000	149.7	414.6	0.65	6.6	480.4
2005 - 2006	38990000	154.4	419.3	0.68	6.8	509.1
2006 - 2007	39680000	156	421.5	0.67	6.9	519.7
2007 - 2008	40550000	155.1	421.7	0.66	6.8	513.8
2008 - 2009	40910000	152.4	419.4	0.64	6.7	495.4
2009 - 2010	40810000	147.4	414	0.63	6.5	464.2

2010 - 2011	41270000	149.5	416.5	0.63	6.7	473.9
2011 - 2012	46930000	155	420.4	0.68	7.0	514
2012 - 2013	48540000	164.6	429.1	0.72	7.4	575.5
2013 - 2014	44880000	155.8	419.5	0.7	7.1	521
2014 - 2015	40350000	156.1	420.9	0.7	7.1	522
2015 - 2016	43770000	154.9	419.5	0.7	7.0	517.1
2016 - 2017	47480000	157.2	423	0.75	7.2	530.7
2017 - 2018	44110000	157.2	423.6	0.75	7.3	526.9
2018 - 2019	44520000	157.5	424.2	0.74	7.4	527.8
2019 - 2020	44020000	159	425.8	0.76	7.6	535.6
2020 - 2021	42480000	167	436.8	0.79	8.1	580.6

Notes: Author's calculations from the MAF-ARF. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY23-ADEP001-008.

Table 7. Distance in Miles of Moves in MAF-ARF Unadjusted Sample (2000-2021)

Year	N	Mean	SD	10th Pct.	Median	90th Pct.
2000 - 2001	38300000	141.9	406.3	0.093	5.25	432.8
2001 - 2002	39650000	138	401	0.082	5.19	409.2
2002 - 2003	39650000	135.3	397	0.09	5.2	392.4
2003 - 2004	40350000	135	396.3	0.1	5.3	390.5
2004 - 2005	41690000	137.9	400	0.11	5.44	410.4
2005 - 2006	43190000	142.8	405.3	0.13	5.69	443.3
2006 - 2007	44510000	142.5	405.2	0.081	5.59	442.8
2007 - 2008	45360000	142.2	406	0.092	5.58	439
2008 - 2009	46100000	138.5	402.2	0.067	5.35	414.8
2009 - 2010	45690000	134.6	397.9	0.081	5.24	388.6
2010 - 2011	47990000	134	397.3	0.036	5.24	377.3
2011 - 2012	53600000	142.6	405.8	0.092	5.98	432.7
2012 - 2013	55070000	154.6	418.9	0.21	6.48	518.3
2013 - 2014	41430000	136.6	394.3	0.058	5.66	399.2
2014 - 2015	45230000	142.8	405.3	0.076	5.71	440.2
2015 - 2016	49210000	141.2	403.3	0.068	5.59	432.9
2016 - 2017	52610000	145.2	409.1	0.13	5.94	458.1
2017 - 2018	49370000	143.7	407.9	0.08	5.91	444.2
2018 - 2019	49980000	143.5	407.9	0.069	5.9	442.3
2019 - 2020	49560000	144.7	409.2	0.065	6.07	448.8
2020 - 2021	47860000	152.1	420.1	0.068	6.43	495.7

Notes: Author's calculations from the MAF-ARF. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY22-ERD002-010.

Summary

The MAF-ARF is a new source of information about geographic mobility in the United States. It is longitudinal and includes annual location information on a much larger portion of the population than cross-sectional or panel surveys. Moreover, it is easily enriched with socio-economic information when linked to other Census Bureau data and administrative records. However, measuring geographic mobility with the MAF-ARF comes with challenges common to research using administrative records. In this paper, we document these data and provide new estimates of geographic mobility. We compare aggregate migration estimates from the MAF-ARF to other published estimates and link the MAF-ARF to survey data at the individual level to document biases in these data.

We utilize the fine geographic detail in the MAF to make novel estimates of move distance and to produce adjusted mobility estimates that can account for some types of measurement error. Our most conservative estimates closely track those from the CPS but are only possible for the years 2000-2011. Our recommended estimates can be made consistently for all available years of data, but in line with other estimates of migration from administrative records, show higher levels of migration than survey estimates. We find more stability in 1-year migration rates than the CPS and find that address changes are twice as common, with around 20% of our sample moving in any year. We find that most moves are relatively short distance, 7 miles or less. Furthermore, the median distance of moves appears to be increasing over the period 2000-2021.

Though migration research using these data should be performed with caution, differences between our estimates and those from other sources is likely in part due to issues

unrelated to data quality. For example, the universes of the MAF-ARF, CPS, ACS and SOI data all differ in some ways. Further, as our comparison to ACS migration reports demonstrates, survey reports of migration are also made with error. Subsequent research using the MAF-ARF should leverage its very large sample, longitudinal capacity, and temporal and spatial detail. However, it should also grapple with issues of sample representativeness, multiple location options before 2012, and potentially upwardly biased estimates of mobility rates.

References

DeWaard, J., Hauer, M., Fussell, E. *et al.* User Beware: Concerning Findings from the Post 2011–2012 U.S. Internal Revenue Service Migration Data. *Popul Res Policy Rev* **41**, 437–448 (2022). https://doi.org/10.1007/s11113-021-09663-6.

Dillon, Michaela. Evaluating the Master Address File – Auxiliary Reference File (MAF-ARF) as a Potential Respondent Retention Source. CARRA Working Paper 18-12. Sept. 2021.

Annemarie Ernsten, David McCollum, Zhiqiang Feng, Dawn Everington & Zengyi Huang (2018) Using linked administrative and census data for migration research, Population Studies, 72:3, 357-367, DOI: 10.1080/00324728.2018.1502463

Keith Finlay and Katie R. Genadek. 2021. Measuring All-Cause Mortality With the Census Numident File. American Journal of Public Health. 111, S141_S148, https://doi.org/10.2105/AJPH.2021.306217.

Brian Foley, Tony Champion and Ian Shuttleworth. How Does Administrative Data Compare with Census Data for the Measurement of Internal Migration? The Case of Northern Ireland, 2001–2011. Applied Spatial Analysis and Policy. April 2020.

Foster, Thomas B., Ellis, Mark and Fiorio, Lee. The Opportunities and Challenges of Linked IRS Administrative and Census Survey Records in the Study of Migration. CARRA Working Paper 2018-06.

Graham, Matthew R., Mark J. Kutsbach and Danielle H. Sandler. Developing a Residence Candidate File for Use With Employer-Employee Matched Data. CES Working Paper Series, 17-40, 2017.

Riordan Frost. Are Americans Stuck in Place? Declining Residential Mobility in the US. Joint Center for Housing Studies of Harvard, May 2020.

Kenneth Johnson, Paul R. Voss, Roger B. Hammer, Glenn V. Fuguitt and Scott McNiven, Temporal and Spatial Variation in Age Specific Net Migration in the United States. Demography, Volume 42-Number 4, November 2005: 791-812.

Raven Molloy and Christopher Smith. U.S. Internal Migration: Recent Patterns and Outstanding Puzzles. Federal Reserve Board Working Paper, Oct. 2019. https://www.bostonfed.org/-/media/Images/research-conference-2019/papers/Session-4-MolloySmith.pdf

U.S. Census Bureau, History: 2000 Census of Population and Housing (Volume 2) U.S. Government Printing Office, Washington, DC, 2009

Wagner, Deborah and Layne, Mary. The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA)

Record Linkage Software. Center for Administrative Records Research and Applications. July 2014.