Analyzing the Residential Status of Demographic Frame Addresses

Brandon P. Pipher, Thomas Mule[1]
U.S. Census Bureau, 4600 Silver Hill Rd., Washington, DC 20233

Abstract:

The Census Bureau's Demographic Frame is a comprehensive, person-level frame consisting of geographic, demographic, social, and economic characteristics. It could operate as a sampling frame for household surveys, reducing respondent burden by using information already available to the federal government, or improving data quality by drawing from the frame for data editing and imputation. It can be used to identify addresses associated with each person and potentially to identify their residence. Person-address records on the Demographic Frame are derived from administrative, third-party, census and survey data records.

A person found on the Demographic Frame may have multiple address records, as they are often associated with several addresses across the various data sources. These multiple address-records create difficulty in placing a person at their correct residential address according to a given reference day. The Demographic Frame has a person-place model process that assigns probabilities to each person-address record and can be used to determine a person's residence on a particular reference date. These models learn from person-place pairs on existing data within the Census Bureau to make predictions about other person-place pairs derived from administrative records.

This analysis will evaluate the accuracy of addresses for people from these models on the Demographic Frame based on a reference date of July 1, 2021. As part of this analysis, we will compare these Demographic Frame addresses to addresses in the 2020 Census and the 2021 American Community Survey frames. This comparison will allow us to examine the Demographic Frame addresses that are not found within these other Census products, which may provide information to help identify whether subsets of these addresses may be more likely to be residential or non-residential addresses. This analysis can potentially improve the quality of the Demographic Frame and its person-place records by increasing the chance of placing a person at their correct residential address.

Keywords: Demographic Frame, Census, American Community Survey, Person-Place Model, Population Estimates, Administrative Records, Sampling Frame

1. Introduction

The result of this report stems from work done within the Continuous Count Study which researches the use and availability of administrative records to generate yearly population estimates in non-census years. The Continuous Count Study also works to quantify the utility and availability of administrative data, and to assess their coverage and quality. As a part of the study, we were researching using the American Community Survey (ACS) and administrative data to generate capture-recapture estimates. This analysis examines how the addresses in the Demographic Frame 2021 Version 1 corresponded with the addresses in the 2021 ACS universe in this study.

[1] This report is released to inform interested parties of (ongoing) research and to encourage discussion (of work in process). Any views expressed on (statistical, methodological, technical, or operational) issues are those of the author(s) and not those of the U.S. Census Bureau. This paper meets all the U.S. Census Bureau's Disclosure Review Board (DRB) standards and has been assigned DRB approval number CBDRB-FY24-0360.

The Demographic Frame (DF) is a person-level frame derived from administrative, third-party, census and survey data records. Using person-place models we can identify housing units associated with each person on the frame for a given reference date. Regarding person-place models, these are models trained using people and their housing unit associations from the Demographic Frame when we have a known housing address for them. The model is then used to estimate the probability of someone living at any of their associated addresses for everyone on the frame. People on the frame are uniquely identified through Protected Identification Keys (PIKs), and housing units are uniquely identified through Master Address File Identifications (MAFIDs) assigned from the Master Address File (MAF). The Master Address is a Census Bureau file that contains an accurate, up to date inventory of all known living quarters in the United States. Some of the administrative records utilized as sources for the Demographic Frame include records from the Internal Revenue Service (IRS), U.S. Department of Housing and Urban Development (HUD), U.S. Postal Service (USPS), and Medicaid.

The American Community Survey is an annual survey conducted by the Census Bureau, and it generates various demographic estimates through a sample of housing units and people. The ACS Sampling universe consists of all residential housing units within the United States. The terminology we will use is ACS-Valid if a housing unit address is within a given ACS Sampling Universe, and ACS-Invalid otherwise. The ACS leverages the MAF to create their sampling universe by filtering out out-of-scope living quarters. By matching between the Demographic Frame and ACS we can identify and analyze people and households that appear on both lists or only one or the other.

2. Design

The Demographic Frame relates multiple addresses to a single person. For the purposes of this analysis, we will arbitrarily de-duplicate the Demographic Frame by assigning each person to a single address based on their highest probability assigned by a boosted trees person-place model. The boosted trees model probabilities are provided by the Frames team within the U.S. Census Bureau. This analysis is conducted only using the household population located within the 50 states and Washington D.C. Following the deduplication process we are left with a frame that has 326 million PIKs with each PIK placed at a single MAFID. We then group these people based on whether their MAFID is found within the ACS Sampling Universe (ACS-Valid) and whether the MAFID was also found on the 2020 Census.

Table 1 shows the people on the Demographic Frame by whether they were also found on the 2021 ACS or the 2020 Census. The 9.2 million people that were found to be on the 2020 Census but at an ACS-Invalid MAFID for the 2021 ACS likely occurs due to the fact the 2021 ACS filter is using a pre-Census MAF extract for their filter. For people at MAFIDs that were invalid for the 2021 ACS and not found on the 2020 Census we also further examined the outcome of the ACS MAF filter to see why it might have been an invalid MAFID for ACS sampling. These 17.8 million people are placed at MAFIDs that are not found on either the Census or the ACS. These MAFIDs could either be housing units missed by the 2020 Census or out-of-scope housing units that we would not want to be assigning people to when using the person-place model, such as non-residential units. A MAFID could fail the ACS filter for multiple reasons, and so we attempted to group them by the most obvious reason it failed. The MAFIDs of these 17.8 million people at MAFIDs that were ACS-Invalid for the 2021 ACS were provided to the Census Bureau's Decennial Frame Branch who analyzed these MAFIDs. These ACS-Invalid MAFIDs were then binned according to the most likely reason they failed the ACS MAF filter according to our most recent data, which at the time of this analysis was the July 2023 MAF extract.

The results of binning Demographic Frame people based on their MAFID's ACS filter status is presented in Table 2. Table 3 provides further details on what the ACS filter groups are meant to represent. The ACS filter frequently leverages information provided by the United States Postal Service through their Delivery Sequence File (DSF) which contains all delivery point addresses serviced by the USPS. The Census is provided with updated data from the DSF every six months.

As shown in Table 2, of the 17.8 million ACS-Invalid people not found on the Census, there were only 3,347 people who were inevitably found to be at a MAFID that is valid for the ACS sampling universe. After examining how many of the 2021 ACS-Invalid MAFIDs remained invalid through 2023 we also looked to see if these MAFIDs were being sourced from administrative records differently than from MAFIDs that were ACS Valid or found by the Census. To examine this, we utilized the Title 26 version of the Demographic Frame. To determine if a PIK-MAFID pair was sourced by a given administrative record we checked to see if it had a last observed date filled for that record type. Note that a PIK-MAFID pair could be sourced from multiple administrative records simultaneously. Table 4 describes the record types that are referenced from the Title 26 version of the Demographic Frame.

3. Results

The results of examining people based on whether their placed DF MAFID was "ACS Valid", "ACS Invalid and on Census", or "ACS Invalid and not on Census" is shown in Table 5. Table 6 shows the same MAFID grouping as Table 5 but further breaks the numbers down based on the administrative record the PIK-MAFID was sourced from. Table 7 shows the same data as Table 6 but is presented as percentages of the total PIKs for a given MAFID grouping and administrative record. Note that our PIK-MAFID pairs could have been sourced from two or more administrative records simultaneously and so the counts and percents presented will not sum together due to a record being found across multiple sources. From Table 7 we can see that 80.8 percent of people at "ACS Invalid and not on Census" MAFIDs are sourced from IRS 1099 records, while 69.0 percent and 69.3 percent of people are at MAFIDs sourced by IRS 1099s for "ACS Valid" or "ACS Invalid and on Census" MAFIDs. Comparing the sourced-by rates for a given administrative record for "ACS Invalid and not on Census" MAFIDs to either "ACS Valid" or "ACS Invalid and on Census" MAFIDs we can see that the PIK-MAFID pairs that are "ACS Invalid and not on Census" are less frequently found amongst all other administrative records. One exception however are IRS 1099 records in which the "ACS Invalid and not on Census" pairs are more frequently sourced from and often the only source these excluded MAFIDs are sourced from.

Additional results are made available in our appendix that further breaks down these numbers by grouping with the ACS Filter status. Table 8 showcases these results as counts and Table 9 showcases them as relative percents. Note again that a single PIK-MAFID pair could be sourced from multiple administrative records at once. If we examine Table 8 we can see that for "ACS Invalid and Non-Census" MAFIDs that failed the ACS Filter due to being non-residential are also found relatively less often amongst all other administrative record sources, with the exception of IRS 1099s in which this is their predominant administrative record source.

4. Conclusions

Following these results, one recommendation would be to exclude MAFIDs on the Demographic Frame that are solely sourced by IRS 1099 sources, except for any MAFIDs that are the sole MAFID for a PIK on the Demographic Frame. Given these ACS-Invalid MAFIDs are often solely sourced from IRS 1099 administrative records, and primarily found to be ACS-Invalid for reasons such as being duplicates or non-residential, we could conclude they are often unreliable MAFIDs to assign PIKs to. Our results also highlight the large number of non-residential and invalid MAFIDs that the demographic frame associates with and is assigning to people. By further investigating these cases we believe we can further improve the quality of the PIK-MAFID cases found on the demographic frame and increase the likelihood of associating a person with their true residential housing unit.

Appendix

Table 1: Demographic Frame People by their MAFID's ACS Valid Status and its presence on the 2020 Census

People on the 2021 Demographic Frame	Count	Percent
2021 ACS Valid	299,152,788	91.7%
2021 ACS Invalid	27,010,992	8.3%
On the 2020 Census	9,242,285	2.8%
Not on the 2020 Census	17,768,707	5.4%
All	326,163,780	100.0%

Table 2: People on the 2021 Demographic Frame People at 2021 Invalid ACS MAFIDs that were not found on the 2020 Census

People on the 2021 Demographic Frame	Count	Percent
2021 ACS Invalid and not on 2020 Census	17,768,707	100.0%
ACS Filter July 2023: Passes ACS Filter	3,347	0.0%
ACS Filter July 2023: Duplicate	2,288,567	12.9%
ACS Filter July 2023: Nonresidential	4,118,055	23.2%
ACS Filter July 2023: Delete	4,353	0.0%
ACS Filter July 2023: Missing from MAF Extract	39	0.0%
ACS Filter July 2023: Invalid DSF	8,874,398	49.9%
ACS Filter July 2023: Date test failure	407,925	2.3%
ACS Filter July 2023: Not in Census / Not on DSF	252,744	1.4%
ACS Filter July 2023: Out-of-scope	175,129	1.0%
ACS Filter July 2023: Purged	1,569,640	8.8%
ACS Filter July 2023: Other	74,510	0.4%

Table 3: ACS Filter Status Dictionary

ACS Filter Group	Description
Passes ACS Filter	Many records that failed the ACS filter prior to July 2021 now pass the filter because of MAF updates from 2020 Census operations as well as updates from the DSF. All other statuses below are for records that did not pass the ACS filter.
Duplicate	Records with either a surviving MAFID or a unit status indicating it is a duplicate or physical merge of some other MAFID.
Nonresidential	Nonresidential records.
Delete	Records deleted by trusted operations after the 2020 Census and records deleted during the 2020 Census that were not later validated by a trusted post-2020 Census source.
Missing from MAF Extract	Records that did not appear on the MAF extract. These are primarily records that were added to the MAF after that MAF extract delivery or were identified as domestic violence shelters which get suppressed from the MAF extract.
Invalid DSF	Records not counted in the 2020 Census that did not qualify as a post-2020 Census because they were listed as non-residential or invalid on the DSF.
Date test failure	Records that did not qualify as post-census DSF adds because they were "old" DSF records.
Not in Census / Not on DSF	Records not counted in the 2020 Census that also did not appear on the DSF and were not validated by a post-census source.
Out-of-scope	Transitory locations, transitory units, or certain types of Group Quarters (GQ) units that are out-of-scope for ACS.
Purged	Records that were physically deleted from the MAF between July 2022 and January 2023.
Other	All other records that fail the ACS filter.

Table 4: Administrative Record Sources for the Demographic Frame

Record Type	Description
CMS MEDB	Centers for Medicare & Medicaid Services (CMS) Medicare Database
CMS TMSIS	Centers for Medicare & Medicaid Services (CMS) Transformed Medicaid Statistical Information System (TMSIS)
HUD FHA	U.S. Department of Housing and Urban Development (HUD) Federal Housing Administration (FHA) Integrated Database
HUD PICTRACS	U.S. Department of Housing and Urban Development (HUD) Public and Indian Housing Information Center (PIC) and Tenant Rental Assistance Certification Systems (TRACS) Longitudinal Dataset
IRS i1040	Internal Revenue Service (IRS) 1040 Taxation Records
IRS i1099	Internal Revenue Service (IRS) 1099 Taxation Records
IRS i1099r	Internal Revenue Service (IRS) 1099r Taxation Records
SSA MBR	Social Security Administration (SSA) Master Beneficiary Record (MBR)
SSA SSR	Social Security Administration (SSA) Supplemental Security Record (SSR)
SSS	Selective Service System (SSS) Registration
STATESWT	State sourced Supplemental Nutrition Assistance Program (SNAP), Women, Infants, and Children (WIC), and Temporary Assistance for Needy Families (TANF) Records
USPS NCOA	U.S. Postal Services (USPS) National Change of Address (NCOA) Records
VSGI	Veteran Service Group of Illinois (VSGI) Consumer Reference Data (CRD)

Table 5: Number of PIKs sourced from a given administrative record based on whether it was 2021 ACS Valid or on the 2020 Census

	ACS Valid	ACS Invalid and on 2020 Census	ACS Invalid and not on 2020 Census
Total PIKs	299,152,788	9,242,285	17,768,707

Table 6: Number of PIKs sourced from a given administrative record based on whether it was 2021 ACS Valid or on the 2020 Census

PIKs sourced by	ACS Valid	ACS Invalid and on 2020 Census	ACS Invalid and not on 2020 Census
CMS MEDB	9,576,243	218,669	198,351
CMS TMSIS	46,223,278	1,416,974	1,080,614
HUD FHA	2,544,856	32,201	13,898
HUD PICTRACS	5,378,788	127,315	97,083
IRS i1040	243,901,787	4,972,695	4,899,956
IRS i1099	206,424,032	6,400,847	14,362,739
IRS i1099r	43,589,614	708,750	784,115
SSA MBR	11,193,517	280,978	245,684
SSA SSR	4,446,232	183,763	149,864
SSS	3,132,069	59,166	45,518
STATESWT	9,709,095	365,561	383,892
USPS NCOA	14,723,547	363,681	260,686
VSGI	21,156,504	541,584	559,158

Table 7: Percent of PIKs sourced from a given administrative record based on whether it was 2021 ACS Valid or on the 2020 Census

PIKs Sourced By	ACS Valid	ACS Invalid and on 2020 Census	ACS Invalid and not on 2020 Census
CMS MEDB	3.2%	2.4%	1.1%
CMS TMSIS	15.5%	15.3%	6.1%
HUD FHA	0.9%	0.3%	0.1%
HUD PICTRAC	1.8%	1.4%	0.5%
IRS i1040	81.5%	53.8%	27.6%
IRS i1099	69.0%	69.3%	80.8%
IRS i1099r	14.6%	7.7%	4.4%
SSA MBR	3.7%	3.0%	1.4%
SSA SSR	1.5%	2.0%	0.8%
SSS	1.0%	0.6%	0.3%
STATESWT	3.2%	4.0%	2.2%
USPS NCOA	4.9%	3.9%	1.5%
VSGI	7.1%	5.9%	3.1%

Table 8: PIKs grouped by their MAFID status and administrative records source
PIK-MAFID pairs can be from multiple sources simultaneously

MAFID Status Groups	Total PIKs	Sourced by CMS MEDB	Sourced by CMS TMSIS	Sourced by HUD FHA	Sourced by HUD PICTRACS	Sourced by IRS i1040	Sourced by IRS i1099	Sourced by IRS i1099r	Sourced by SSA MBR	Sourced by SSA SSR	Sourced by SSS	Sourced by STATESWT	Sourced by USPS NCOA	Sourced by VSGI
ACS 2021 Valid	299,152,788	9,576,243	46,223,278	2,544,856	5,378,788	243,901,787	206,424,032	43,589,614	11,193,517	4,446,232	3,132,069	9,709,095	14,723,547	21,156,504
ACS 2021 Invalid and on 2020 Census	9,242,285	218,669	1,416,974	32,201	127,315	4,972,695	6,400,847	708,750	280,978	183,763	59,166	365,561	363,681	541,584
ACS 2021 Invalid and not on 2020 Census	17,768,707	198,351	1,080,614	13,898	97,083	4,899,956	14,362,739	784,115	245,684	149,864	45,518	383,892	260,686	559,158
ACS Filter July 2023: Date test failure	407,925	13,360	61,663	1,259	5,854	252,195	263,218	36,728	19,286	8,052	3,235	16,867	25,300	30,031
ACS Filter July 2023: Delete	4,353	85	487	0	72	2,787	3,029	394	164	85	9	106	130	183
ACS Filter July 2023: Duplicate	2,288,567	41,861	251,247	5,674	35,408	1,117,976	1,676,960	184,092	52,607	32,354	13,234	84,579	70,152	110,716
ACS Filter July 2023: Invalid DSF	8,874,398	50,425	351,261	3,488	30,145	1,454,408	7,655,203	255,412	61,548	59,895	14,260	141,735	55,563	247,823
ACS Filter July 2023: Missing from MAF Extract	39	34	0	3	0	0	29	36	10	0	0	26	4	21
ACS Filter July 2023: Nonresidential	4,118,055	50,264	216,835	475	10,191	1,086,158	3,270,477	139,549	61,213	25,377	6,561	54,513	70,869	111,929
ACS Filter July 2023: Not in Census / Not DSF	252,744	3,611	28,305	789	2,343	113,542	182,664	16,579	4,405	3,426	1,131	9,989	3,622	9,013
ACS Filter July 2023: Other	74,510	1,541	13,771	249	361	14,460	52,075	2,493	1,494	2,470	120	6,129	1,016	2,030
ACS Filter July 2023: Out-of-scope	175,129	4,409	12,728	48	1,225	68,072	129,758	8,416	4,316	3,260	533	6,631	5,755	8,949
ACS Filter July 2023: Passes ACS Filter	3,347	87	454	2	93	1,555	2,038	167	192	150	0	55	169	229
ACS Filter July 2023: Purged	1,569,640	32,674	143,863	1,911	11,391	788,803	1,127,288	140,249	40,449	14,795	6,435	63,262	28,106	38,234

Table 9: Percent of PIKs grouped by their MAFID status and administrative records source
PIK-MAFID pairs can be from multiple sources simultaneously

MAFID Status Groups	Total PIKs	Sourced by CMS MEDB	Sourced by CMS TMSIS	Sourced by HUD FHA	Sourced by HUD PICTRACS	Sourced by IRS i1040	Sourced by IRS i1099	Sourced by IRS i1099r	Sourced by SSA MBR	Sourced by SSA SSR	Sourced by SSS	Sourced by STATESWT	Sourced by USPS NCOA	Sourced by VSGI
ACS 2021 Valid	299,152,788	3.2%	15.5%	0.9%	1.8%	81.5%	69.0%	14.6%	3.7%	1.5%	1.0%	3.2%	4.9%	7.1%
ACS 2021 Invalid and on 2020 Census	9,242,285	2.4%	15.3%	0.3%	1.4%	53.8%	69.3%	7.7%	3.0%	2.0%	0.6%	4.0%	3.9%	5.9%
ACS 2021 Invalid and not on 2020 Census	17,768,707	1.1%	6.1%	0.1%	0.5%	27.6%	80.8%	4.4%	1.4%	0.8%	0.3%	2.2%	1.5%	3.1%
ACS Filter July 2023: Date test failure	407,925	3.3%	15.1%	0.3%	1.4%	61.8%	64.5%	9.0%	4.7%	2.0%	0.8%	4.1%	6.2%	7.4%
ACS Filter July 2023: Delete	4,353	2.0%	11.2%	0.0%	1.7%	64.0%	69.6%	9.1%	3.8%	2.0%	0.2%	2.4%	3.0%	4.2%
ACS Filter July 2023: Duplicate	2,288,567	1.8%	11.0%	0.2%	1.5%	48.9%	73.3%	8.0%	2.3%	1.4%	0.6%	3.7%	3.1%	4.8%
ACS Filter July 2023: Invalid DSF	8,874,398	0.6%	4.0%	0.0%	0.3%	16.4%	86.3%	2.9%	0.7%	0.7%	0.2%	1.6%	0.6%	2.8%
ACS Filter July 2023: Missing from MAF Extract	39	87.2%	0.0%	7.7%	0.0%	0.0%	74.4%	92.3%	25.6%	0.0%	0.0%	66.7%	10.3%	53.8%
ACS Filter July 2023: Nonresidential	4,118,055	1.2%	5.3%	0.0%	0.2%	26.4%	79.4%	3.4%	1.5%	0.6%	0.2%	1.3%	1.7%	2.7%
ACS Filter July 2023: Not in Census / Not DSF	252,744	1.4%	11.2%	0.3%	0.9%	44.9%	72.3%	6.6%	1.7%	1.4%	0.4%	4.0%	1.4%	3.6%
ACS Filter July 2023: Other	74,510	2.1%	18.5%	0.3%	0.5%	19.4%	69.9%	3.3%	2.0%	3.3%	0.2%	8.2%	1.4%	2.7%
ACS Filter July 2023: Out-of-scope	175,129	2.5%	7.3%	0.0%	0.7%	38.9%	74.1%	4.8%	2.5%	1.9%	0.3%	3.8%	3.3%	5.1%
ACS Filter July 2023: Passes ACS Filter	3,347	2.6%	13.6%	0.1%	2.8%	46.5%	60.9%	5.0%	5.7%	4.5%	0.0%	1.6%	5.0%	6.8%
ACS Filter July 2023: Purged	1,569,640	2.1%	9.2%	0.1%	0.7%	50.3%	71.8%	8.9%	2.6%	0.9%	0.4%	4.0%	1.8%	2.4%