Learning More About HUD-Assisted Tenants Through Data Linkage

Shawn Bucholtz
U.S. Department of Housing and Urban Development

Emily Molfino
U.S. Census Bureau

Quentin Brummet
NORC at the University of Chicago

The views expressed in this article are those of the author and do not represent the official positions or policies of the Office of Policy Development and Research, the U.S. Department of Housing and Urban Development, or the U.S. Government.

1. Introduction

The U.S. Department of Housing and Urban Development (HUD) administers several rental assistance programs that help low-income households afford their rental units, aiding demographics such as seniors, people with disabilities, and veterans. These programs include the Public Housing (PH), Housing Choice Voucher (HCV), and project-based rental assistance (PBRA) programs.[1] Altogether, HUD's rental assistance programs provide housing for over 4.6 million households or about 10.5 percent of all households in the United States.

To administer rental assistance programs in a manner consistent with statutory, regulatory, and program-specific requirements, HUD must collect information from its beneficiaries. Like many federal programs, however, HUD's information collection is generally limited to the information necessary to implement the program—a legal requirement stemming from the Paperwork Reduction Act of 1995.[2] The limited information collected by an agency is often not sufficient to fully monitor ongoing program performance or evaluate longer-term program impact, including both impacts on the beneficiaries themselves and the public.

Linking administrative records to existing surveys provides a promising method for low-cost evaluation of program performance and impact, and HUD has been a leader in this area. As one example, HUD rental assistance administrative records were linked to the Center for Disease Control's National Health and Nutrition Examination Survey and National Health Interview Survey. Researchers using this linked data have produced several important findings about HUD-assisted households, including findings about blood lead levels (Ahrens et al., 2016); cigarette smoking (Wang et al., 2015); levels of physical activity (Wong et al., 2018); health

---

[1] For the purposes of this article, other small project-based programs are included in the PBRA total.
[2] U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Assistance, Paperwork Reduction Act (PRA), 44 U.S.C. § 3501 *et seq*.

insurance uptake (Simon et al., 2017); health services use (Brucker, Helms, and Souza, 2018); and overall adult health (Fenelon et al., 2017).

Numerous other researchers have used the U.S. Census Bureau's linking infrastructure to link administrative records to existing surveys to study a range of topics, including the impact of Medicaid expansion on mortality (Miller et al., 2019); family relationships (O'Hara, Shattuck, and Goerge, 2017); economic inequality and mobility (Medalia et al., 2019); minority-owned and women-owned businesses (Jarmin, Krizan, and Luque, 2016); the effect of pollution exposure on adult wages, education attainment, and incarceration (Voorheis, 2017); and the impact of transportation on physical and mental health and the environment (Cavoli et al., 2015).

Following these numerous examples, social and data scientists at HUD and the Census Bureau developed a procedure to link HUD rental assistance administrative records to the American Community Survey (ACS), thereby identifying ACS households as receiving HUD rental assistance. ACS contains a wealth of household and demographic information that is not currently collected by HUD. Some examples of information in ACS include—

- Type of occupation and commuting mode.
- Veteran status.
- Health insurance status.
- Expanded racial categories and household relationship types.
- Internet access.

This new linked dataset allows HUD to gain insights into HUD-assisted housing units and households that would otherwise not be possible with current rental assistance administrative records, potentially leading to more robust program evaluation. Our record linkage process takes full advantage of both address- and person-level matching to overcome HUD address quality issues. A full technical description of the data sources, linking process, and linking performance metrics is available in Bucholtz, Molfino, and Brummet (2020).

The remainder of this article describes the data and record linkage process, presents an evaluation of the linkage quality, and discusses early insights gleaned from the linked data.[3] We conclude with a discussion of how to access the linked data for further research.

2. Data

We linked the HUD rental housing assistance administrative records to ACS. To accomplish this, we relied on two other Census Bureau data sources integral to the linking process: the Master Address File (MAF) and the Social Security Administration (SSA) Numerical Identification (Numident) File. A fuller description of the data sources is available in Bucholtz, Molfino, and Brummet (2020).

ACS internal use file (IUF) microdata differ from the public versions of the ACS microdata in three important ways. First, the ACS IUFs contain precise location information

---

[3] The U.S. Census Bureau reviewed all tables in this article for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release.

(that is, housing unit address), which can be used to link to other data sources through address matching or other spatial matching techniques. Second, the ACS IUFs contain respondent names, which can be used to link to other data sources through person-matching techniques. Finally, the ACS IUFs include all the survey responses, whereas the ACS public use microdata typically includes about two-thirds of the actual number of respondents (U.S. Census Bureau, 2009).

The HUD rental assistance administrative records are derived from HUD forms 50058 (for PH and HCV) and 50059 (for PBRA). These forms are used to collect information from tenants who are participating or seeking to participate in rental assistance programs. Although the data collection forms are paper, virtually all PHAs implement an electronic version of the forms. The electronic data are transferred to HUD daily, and HUD uses them to monitor program compliance and performance.

The Census Bureau's MAF is a collection of all addresses in the United States. MAF was originally built for the 2000 Decennial Census, using addresses from the 1990 Decennial Census as well as the U.S. Postal Service (USPS) Delivery Sequence File (DSF) (National Research Council, 2004). Currently, the MAF is updated twice a year using the USPS DSF and other USPS information, as well as information gathered during other census surveys and decennial preparation operations (U.S. Census Bureau, 2014). A key feature of MAF is that every address it contains is assigned a unique key called a MAF identifier, or MAFID.

The SSA Numident file contains the name, date, place of birth, and parent's names for each social security number (SSN). It also contains all transactions for an SSN, such as name changes and deaths. Using the SSA's Numident file as a base, the Census Bureau builds their own version of a Numident file regularly. This process augments the SSA's Numident with information from other federal and state administrative records to include current address and household composition (Wagner and Layne, 2014).

3. Record Linkage Process

At a high level, the record linkage process included four steps, which are outlined below. Some steps of the linkage process are considered probabilistic, meaning a linkage between two records is considered "acceptable" if a "probability of linkage" threshold is achieved. Other steps are considered deterministic, meaning the linkage is either a "yes" or "no." A complete description of the record linkage process, as well as potential problems with the linkage process, is available in Bucholtz, Molfino, and Brummet (2020).

The first step of the linking process is to extract HUD rental assistance administrative records from HUD's systems. One extract is made per year and represents all households receiving HUD assistance at any time during the calendar year.

The second step in the linking process is address matching. This step begins by "cleaning" the addresses in HUD rental assistance administrative records. The Census Bureau uses standardization algorithms that edit and standardize the addresses. The standardization

algorithms are themselves probabilistic.[4] Then, the HUD rental assistance administrative record addresses are probabilistically matched to the Census Bureau's MAF addresses using a Census Bureau address matching algorithm. The result is that each HUD rental assistance administrative record receives a unique MAFID based on its address. Finally, we deterministically match the HUD rental assistance administrative records to ACS using MAFID. It is important to note that each ACS record already had a MAFID because the ACS housing unit sample is derived from housing unit addresses in MAF.

The third step in the linking process is person-level matching. This step starts with a process to deterministically match HUD rental assistance administrative records (which include SSNs) to the Numident file using SSNs. This results in each HUD rental assistance administrative record receiving a "PIK or Protected Identification Keys," which is functionally a pseudo-SSN used to safeguard real SSNs throughout the matching process. Then, we probabilistically match the ACS households to the Numident file. The Census Bureau developed an algorithm to match ACS respondents to the Census Bureau's Numident file using the information provided by the ACS respondent. This algorithm is embedded in a system called the Person Identification Validation System, or PVS. As discussed in Wagner and Layne (2014), PVS uses the respondent's name, gender, address, date of birth, and household relationship to match a respondent to the Numident file probabilistically. This step results in each ACS person record receiving a PIK value. Finally, we deterministically matched HUD rental assistance administrative records to ACS using the PIK values.

The fourth and final step in the linkage process is to make a final linkage determination. In our linking process, any ACS housing unit that matches to a HUD rental assistance administrative record by a MAF address match is considered a valid link. Any additional ACS household that matches to a HUD rental assistance administrative record by SSN match is also considered a valid link, so long as the HUD administrative record and the ACS record are in the same county.

Exhibit 1 shows the number of ACS records linked to a HUD administrative record by the type of link. Although not the subject of this article, we speculate that the downward trend in the total number of ACS records linked to a HUD administrative record reflects a general downward trend in response rates for HUD-assisted households. The authors have observed a similar trend in another household survey, the American Housing Survey.[5]

Exhibit 1. Breakdown of American Community Survey/HUD links made by Address (MAF) and Person (PIK)

	Year						
	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**
MAF-Matched	60,000	63,500	55,500	58,000	57,000	54,500	51,500
	81.6%	81.9%	82.2%	82.9%	82.6%	83.8%	84.4%
PIK-Matched	13,500	14,000	12,000	12,000	12,000	10,500	9,500

---

[4] Refer to Brummet (2015) for further descriptions of this process.
[5] Based on unpublished analysis of 2015, 2017, and 2019 American Housing Survey internal use files.

| | 18.4% | 18.1% | 17.8% | 17.1% | 17.4% | 16.2% | 15.6% |
| **Total** | 73,500 | 77,500 | 67,500 | 70,000 | 69,000 | 65,000 | 61,000 |

MAF = Master Address File. PIK = Protected Identification Key.
Note: Numbers are rounded to comply with Census Bureau disclosure guidelines.

Finally, it is important to note that the record linkage process is not perfect. There are numerous ways in which the process can fail to link an ACS and HUD record (false negative) or incorrectly link an ACS and HUD record when they do not represent the same housing unit (false positive). For instance, the addresses in HUD's rental assistance administrative records may not be of sufficient quality to be cleaned (step 2) or uniquely matched to a MAF address (step 2). Likewise, the address matching algorithm (step 2) may fail. The person-matching algorithm (step 3) used to match ACS household members to the Numident file can produce both false positives and false negatives, as persons could move across units, and there is a small amount of error inherent in the PVS process (Layne, Wagner, and Rothhaas, 2014).

4. Record Linkage Quality Assessment

To determine whether the multifaceted linking process performed well, we compared the "pre-linking" count of HUD rental assistance administrative records with the "post-linking" ACS weighted estimate of ACS housing units identified as HUD-assisted. All else being equal, if the linking process performs well, the post-link ACS weighted estimate of HUD-assisted units should be very similar to the pre-link known record count.

Exhibit 2 below presents linking quality metrics for 2015 through 2017. The table shows that HUD provided the Census Bureau with 4.74 million rental assistance administrative records housing unit records in 2017. When these records were linked to ACS housing units, the weighted estimate of HUD-assisted housing units was 4.62 million, or 97.3 percent of the real total. Across all years of the data linkage (2011–2017), the ACS-weighted estimate of HUD-assisted housing units ranges from 97.0 to 99.4 percent. There is some variation in linkage rate across the three categories of HUD programs, however, with the PBRA program consistently performing worse than PH or HCV. Reasons for this difference are explored in Bucholtz, Molfino, and Brummet (2020), but are generally due to variations in the quality of HUD addresses.

Given the results in exhibit 2, a reasonable conclusion is that the linking process performed well enough to ensure that the ACS housing units flagged as HUD-assisted units are a representative cross-section of all possible ACS housing units that are truly HUD-assisted units. In statistical terms, although there are false negatives (positives), they appear to be limited in quantity, and we feel their omission (inclusion) should not result in biased estimates of housing or household characteristics of HUD-assisted households. Regardless of the extent to which our linking process introduced any bias, a method for overcoming this bias is described in Bucholtz, Molfino, and Brummet (2020).

Exhibit 2. Results of American Community Survey/HUD Administrative Linking

		All	PH	HCV	PBRA
2015	HUD records provided to Census	4,757,000	998,200	2,265,000	1,494,000
	ACS estimate of HUD-assisted households	4,678,000	1,021,000	2,256,000	1,400,000
	ACS estimate as share of HUD records	98.3%	102.3%	99.6%	93.7%
	ACS 90% Margin of Error	0.7%	1.4%	1.2%	1.3%
2016	HUD records provided to Census	4,760,000	1,014,000	2,300,000	1,446,000
	ACS estimate of HUD-assisted households	4,623,000	1,001,000	2,248,000	1,374,000
	ACS estimate as share of HUD records	97.1%	98.7%	97.7%	95.0%
	ACS 90% Margin of Error	0.7%	1.5%	1.1%	1.1%
2017	HUD records provided to Census	4,744,000	977,100	2,313,000	1,453,000
	ACS estimate of HUD-assisted households	4,615,000	979,700	2,268,000	1,367,000
	ACS estimate as share of HUD records	97.3%	100.3%	98.1%	94.1%
	ACS 90% Margin of Error	0.8%	1.5%	1.1%	1.3%

ACS = American Community Survey. HCV = Housing Choice Voucher program. PH = public housing. PBRA = project-based rental assistance.

5. Uses of the Linked Data

In this section, we illustrate two uses of the linked data to produce estimates that are otherwise not feasible to derive using HUD rental assistance administrative records alone.

On HUD forms 50058 and 50059, current and prospective HUD-assisted renters supply a host of demographic information, including age, sex, race, and ethnicity. Consistent with federal guidelines governing the collection of race and ethnicity data, HUD collects race information using five standard categories: White, Black or African-American, Asian-American, American Indian or Alaska Native, and Native Hawaiian or Pacific Islander. ACS follows the same federal guidelines but expands the number of categories for Asian-American from one (Asian American) to six detailed Asian race categories and offers respondents a write-in option.

The linked ACS/HUD data can be used to estimate the number of HUD-assisted householders reporting their race as Asian, by detailed Asian race category. Exhibit 3 below presents these results from the 2013–2017 ACS 5-year data. The results reveal significant variation within the Asian race groups in the share of households receiving HUD assistance relative to those *eligible* for HUD assistance. While it is beyond the scope of the article to

explain these differences further, this analysis illustrates the potential value of the data linkage for better understanding who is served by HUD rental assistance programs.

Exhibit 3. Detailed Asian Race for HUD-Assisted Households, 2013–2017

Householder Race	HUD-Assisted Households	Households Eligible for HUD Assistance	Share of Eligible Households Receiving HUD Assistance (%)
Asian Indian	5,973	105,533	6
Cambodian	5,480	16,640	33
Chinese	53,810	259,710	21
Filipino	12,340	82,330	15
Hmong	4,515	14,389	31
Japanese	3,117	34,187	9
Korean	24,000	116,480	21
Laotian	2,289	9,847	23
Other Asian or Two Groups	9,879	116,379	8
Vietnamese	32,370	89,910	36

As another example, HUD and the U.S. Department of Veterans Affairs (VA) partner to implement the VA Supportive Housing (VASH) program, which provides housing vouchers to homeless veterans. As of 2017, the HUD-VASH program provided housing to nearly 88,000 households with a veteran (Montgomery and Cusack, 2017). HUD leaders long suspected that other HUD rental assistance programs provided housing to many additional veterans that were not part of the VASH program. As is the case with detailed race and ethnicity data, however, HUD forms 50058 and 50059 do not collect information on veteran's status.

The linked ACS/HUD data can be used to estimate the number of HUD-assisted households with a veteran to inform this program. Exhibit 4 below presents these results by year from the 1-year ACS for 2011 through 2017. The results reveal that HUD is serving between 285,000 and 324,000 households with a veteran.

Exhibit 4. Number of HUD-Assisted Households with Veterans by Year, 2011–2017

Year	HUD-Assisted Households with a Veteran
2011	302,000
2012	314,200
2013	289,900
2014	288,200
2015	286,000
2016	285,600
2017	291,900

Source: 2011–2017 American Community Survey, 1-year data

6. Conclusion

Using a multifaceted approach, we have linked administrative data from HUD's rental housing assistance programs to ACS housing units using address and SSN information for years 2011 through 2017. In the future, this work will continue, and we plan to link HUD administrative records to future years of the ACS as they become available. In each year of the ACS, we identify 61,000 to 78,000 ACS households as being HUD-assisted. Our analysis of the data linkage quality suggests that false-positive links and false-negative links are minimal, enabling high-quality analysis of the linked data. A full technical report on the process is found in Bucholtz, Molfino, and Brummet (2020).

Our goal with this project was to develop the linkage process and build the linked data sets so researchers can further explore the data. Access to the ACS/HUD IUFs is available to researchers through a Federal Statistical Research Data Center (FSRDC), after obtaining Special Sworn Status and approval for their project. The process is as follows:

1. Identify the FSRDC located nearest to you (https://www.census.gov/fsrdc).
2. Contact the FSRDC administrator to explain your interest in using the linked ACS/HUD IUF.
3. Complete a proposal following the proposal guidelines (https://www.census.gov/ces/pdf/Research_Proposal_Guidelines.pdf).
4. If your proposal is approved, complete additional application materials and submit to a background check.
5. Conduct your research at the FSRDC.
6. Submit your results to the FSRDC for clearance.

Authors
Shawn Bucholtz is the Head Statistical Official and Director of Housing and Demographic Analysis in the Office of Policy Development and Research, U.S. Department of Housing and Urban Development.

Emily Molfino is a Survey Statistician in the Associate Directorate for Demographic Programs at the U.S. Census Bureau.

Quentin Brummet is a Senior Research Methodologist in the Statistics and Methodology department at NORC at the University of Chicago.

References
Ahrens, Katherine A., Barbara A. Haley, Lauren M. Rossen, Patricia C. Lloyd, and Yutaka Aoki. 2016. "Housing Assistance and Blood Lead Levels: Children in the United States, 2005–2012," *American Journal of Public Health* 106(11): 2049–2056. doi: 10.2105/AJPH.2016.303432.

Brucker, Deborah L., Veronica Helms, and Teresa Souza. 2018. "Health and Health Services Access Among Adults With Disabilities Who Receive Federal Housing Assistance," *Housing Policy Debate* 2: 248–266. doi: 10.1080/10511482.2017.1357048.

Brummet, Quentin. 2015. "Matching Addresses between Household Surveys and Commercial Data. 4. Center for Administrative Records Research and Applications." Working paper #2015-04 (July 6). Washington, D.C.: U.S. Census Bureau, Center for Administrative Records Research and Applications.

Bucholtz, Shawn, Emily Molfino, and Quentin Brummet. 2020. *Identifying Subsidized Housing Units within the American Community Survey through Administrative Record Linkage: A Technical Report.* Washington, D.C.: U.S. Department of Housing and Urban Development.

Cavoli, Clemence, Nicola Christie, Jennifer Mindell, and Helena Titheridge. 2015. "Linking Transport, Health and Sustainability: Better Data Sets for Better Policy-Making," *Journal of Transport and Health* 2(2): 111–119. doi: 10.1016/j.jth.2014.08.001.

Fenelon, Andrew, Patrick Mayne, Alan E. Simon, Lauren M. Rossen, Veronica Helms, Patricia Lloyd, Jon Sperling, Barry L. Steffen. 2017. "Housing Assistance Programs and Adult Health in the United States," *American Journal of Public Health* 107(4): 571–578. doi: 10.2105/ajph.2016.303649.

Jarmin, Ron, C.J. Krizan, and Adela Luque. 2016. "Small Business Growth and Failure During the Great Recession: The Role of House Prices, Race and Gender." Working paper 2016-08. Washington, D.C.: U.S. Census Bureau, Center for Administrative Records Research and Applications.

Layne, Mary, Deborah Wagner, and Cynthia Rothhaas. 2014. "Estimating Record Linkage False Match Rate for the Person Identification Validation System." Working paper #2014-02 (July 1). Washington, D.C.: U.S. Census Bureau, Center for Administrative Records Research and Applications.

Medalia, Carla, Bruce D. Meyer, Amy B. O'Hara, and Derek Wu. 2019. "Linking Survey and Administrative Data to Measure Income, Inequality, and Mobility," *International Journal of Population Data Science* 4(1). doi: 10.23889/ijpds.v4i1.939.

Miller, Sarah, Sean Altekruse, Norman Johnson, and Laura R. Wherry. 2019. Medicaid and Mortality: New Evidence from Linked Survey and Administrative Data. NBER Working Paper No. 26081. Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w26081.

Montgomery, Ann Elizabeth, and Meagan Cusack. 2017. *HUD-VASH Exit Study: Final Report.* Washington, D.C.: U.S. Department of Housing and Urban Development.

National Research Council. 2004. *The 2000 Census: Counting Under Adversity.* Washington, D.C.: The National Academies Press.

O'Hara, Amy, Rachel M. Shattuck, and Robert M. Goerge. 2017. "Linking Federal Surveys with Administrative Data to Improve Research on Families," *The ANNALS of the American Academy of Political and Social Science* 669(1): 63–74. 10.1177/0002716216678391.

Simon, Alan E., Andrew Fenelon, Veronica Helms, Patricia C. Lloyd, and Lauren M. Rossen. 2017. "HUD Housing Assistance Associated With Lower Uninsurance Rates And Unmet Medical Need," *Health Affairs (Millwood)* 36(6): 1016–1023. doi: 10.1377/hlthaff.2016.1152.

U.S. Census Bureau. 2014a. *American Community Survey Design and Methodology: Chapter 7 Data Collection and Capture for Housing Units.* Washington, D.C.: U.S. Census Bureau.

———. 2014b. *American Community Survey Design and Methodology: Chapter 3 Frame Development.* Washington, D.C.: U.S. Census Bureau.

———. 2009. *A Compass for Understanding and Using American Community Survey Data: What Users of Data for American Indians and Alaska Natives Need to Know.* Washington D.C.: U.S. Census Bureau, U.S. Government Printing Office.

Voorheis, John. 2017. "Air Quality, Human Capital Formation and the Long-Term Effects of Environmental Inequality at Birth." Working paper 2017-05. U.S. Census Bureau, Center for Administrative Records Research and Applications.

Wagner, Deborah, and Mary Layne. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software." Working paper 2014-01. Washington, D.C.: U.S. Census Bureau, Center for Administrative Records Research and Applications.

Wang, Teresa W., Veronica E. Helms, Peter J. Ashley, and Brian A. King. 2018. "Characteristics and Correlates of Cigarette Smoking Status Among U.S. Adults Receiving Federal Housing Assistance," *Preventing Chronic Disease* 15: E48. doi: 10.5888/pcd15.170395.

Wong, Michelle S., Eric T. Roberts, Carolyn M. Arnold, and Craig E. Pollack. 2018. "HUD Housing Assistance and Levels of Physical Activity Among Low-Income Adults," *Preventing Chronic Disease* 15: E94. doi: 10.5888/pcd15.170517.