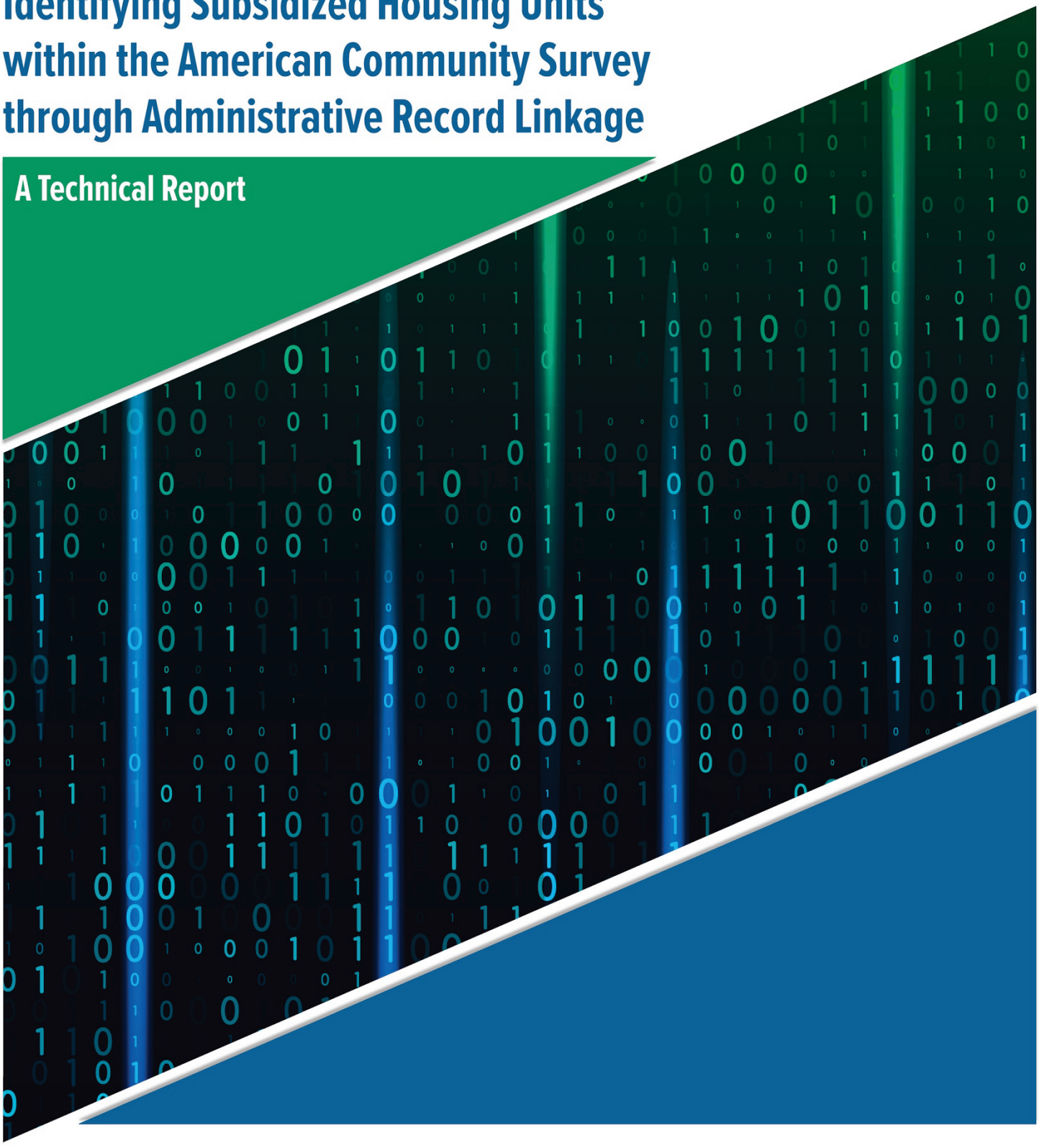
Identifying Subsidized Housing Units within the American Community Survey through Administrative Record Linkage

A Technical Report



U.S. Department of Housing and Urban Development | Office of Policy Development and Research

Identifying Subsidized Housing Units Within the American Community Survey Through Administrative Record Linkage: A Technical Report

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

Emily Molfino, U.S. Census Bureau

Quentin Brummet, National Opinion Research Center

Abstract

In this report, we describe the development of a multifaceted record linkage approach to identify federally subsidized rental housing units in the U.S. Census Bureau's American Community Survey (ACS). Administrative data from the U.S. Department of Housing and Urban Development's (HUD's) rental housing assistance programs are linked to ACS housing units using a combination of three linkage algorithms. The first linkage algorithm is traditional address matching and is used for HUD rental assistance administrative records with "clean" addresses. The second linkage algorithm, which we call "imputed address" matching, is used when HUD's rental assistance administrative records have incomplete addresses. This algorithm relies on the Census Bureau's Master Address File (MAF) as a source for imputing missing apartment numbers in HUD records. The third linkage algorithm is a person-based linkage that relies on the Census Bureau's Numerical Identification (Numident) file and the Census Bureau's Person Identification Validation System (PVS), which uses a compilation of administrative records from various federal sources. When we used our multifaceted record linkage approach, our annual linked data sets resulted in an estimated total number of HUD-assisted rental units between 98 and 100 percent of the known number receiving HUD rental assistance.

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. These households include those with seniors, disabled persons, and veterans. The largest of these programs is the Housing Choice Voucher (HCV) program, with approximately 2.3 million households receiving rental assistance. The second largest of these programs is project-based rental assistance (PBRA), with approximately 1.3 million households receiving rental assistance; the third largest program, public housing (PH), currently provides housing for approximately 950,000 households.[1] These three programs and myriad smaller HUD rental assistance programs provide housing for more than 4.6 million households, or about 3.8 percent of all households in the United States. We hereinafter refer to these housing units, and the households occupying these units, as "HUD-assisted."

To administer rental assistance programs in a manner consistent with statutory, regulatory, and program-specific requirements, HUD must collect information from the beneficiaries. Like many federal programs, however, HUD's information collection is generally limited to only the information necessary to implement the program—a legal requirement stemming from the Paperwork Reduction Act of 1995 (U.S. Department of Justice, n.d.).

It is often the case that the limited information collected by an agency is not sufficient to fully monitor ongoing program performance or evaluate longer term program effects, including the effects on the beneficiaries themselves and the public through positive and negative spillovers. Evaluating program effects often requires additional surveys, which are expensive. This shortcoming in the ability to evaluate programmatic impact is well known to policymakers and members of research and advocacy communities. The 2016 Evidence-based Policymaking Commission Act established a commission charged with making recommendations to improve the federal government's ability to monitor programs. Some of the Commission's recommendations were put into law with the Foundations for Evidence-based Policymaking Act of 2019. Among the many aspects of the new law is a requirement for agencies to develop written evaluation plans and establish evaluation officers (U.S. Commission on Evidence-based Policymaking, 2017).

A promising method for low-cost evaluation of program performance and impact is to link beneficiary records, commonly referred to as administrative records, to other administrative data sources that measure outcomes for program beneficiaries. An impactful recent example of this is by Raj Chetty, Nathan Hendren, and Lawrence Katz (2015). The authors matched HUD administrative data from the Moving to Opportunity experiment to tax data from the Internal Revenue Service (IRS) to study long-term outcomes for children living in HUD-assisted housing. Among several findings, they found that young children moving to lower poverty neighborhoods increased their future college attendance and earnings.

[1]HUD administers numerous other rental housing finance and loan programs; however, this article pertains only to matching administrative records from the rental assistance programs.

Another promising method for low-cost evaluation of program performance and effect is to link administrative records to existing surveys. 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 (NHANES) and National Health Interview Survey (NHIS). Researchers using the linked data have produced several important findings about HUD-assisted households. These findings center around blood lead levels (Ahrens et al., 2016); cigarette smoking (Wang et al., 2015); levels of physical activity (Wong et al., 2018); health insurance uptake (Simon et al., 2017); health services use (Brucker, Helms, and Souza, 2017); and overall adult health (Fenelon et al., 2017).

Numerous other researchers have linked 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).

It is in the spirit of the second method (linking administrative records to existing surveys) that social scientists and data scientists at HUD and the Census Bureau linked HUD-assisted housing unit and household administrative records to American Community Survey (ACS) housing unit records, thereby identifying ACS households as receiving HUD rental assistance. The ACS contains a wealth of household and demographic information that is not currently collected by HUD. Some examples of information in the 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 administrative data, potentially leading to more robust program evaluation. For instance, with these matched data, HUD now knows that approximately 290,000 veterans live in HUD-assisted housing, and approximately 55,000 HUD-assisted households include a person of Chinese descent.

The remainder of this article describes the technical approach used to link HUD administrative records to the ACS and describes some additional insights gleaned from the linked data. Section 2 presents an important note about the process we used to identify HUD-assisted housing units in the ACS. Section 3 describes the two data sources being linked (ACS and HUD rental housing assistance administrative records) and two Census Bureau data sources that are integral to the record linkage process. Section 4 provides a step-by-step technical explanation of the record linkage process. Section 5 presents a qualitative discussion of data linkage quality and potential linkage failures. Section 6 offers a numerical evaluation of the linkage quality. Section 7

describes a technique to overcome potential bias in the linked data due to the record linkage process. Section 8 presents a conclusion and describes how to access the linked data.

2. Two Important Points About the Record Linkage Process

Readers should consider two important and closely related points as they dive into the technical aspects of the record linkage process described in this report. The two points are described below.

2.1 Why We Used Address Matching and Person-level Matching

Within the Census Bureau, two of the most common techniques for linking data sources are address- and person-level matching. These techniques have a rich history of application across many different types of data. Nonetheless, these two linking processes are not perfect. Patterns of linkage rates between administrative data and survey data have been shown to fluctuate depending on person-level characteristics (Brummet et al., 2018; Luque and Bhaskar, 2014; Rastogi et al., 2012). Census Bureau staff have also found that address-based record linkage rates vary by housing type—an issue due mostly to poor address quality in one or more of the linked data sets (Clark et al., 2018).[2] Because such linkage errors do not occur randomly, they increase the risks that resulting estimates based on linked records are biased.

Going into this project, we were well aware of two data quality issues. First, HUD address quality was not perfect (as described in Section 3.2). Because of this, we knew that ACS-HUD record linkage based only on address matching would fail to identify all HUD-assisted housing units with the ACS samples (as described in Sections 5.1 and 5.2). Second, the Census Bureau's process for matching administrative records to surveys using information about people (names, gender, and so on) is good but not perfect (as described in Section 5.3). We knew that record linkage based only on person-level matching would fail to identify all HUD-assisted housing units in the ACS samples. In addition, we knew that record linkage failure, whether from address-matching failures or person-level matching failures, can introduce bias into an analysis of linked data.

To reduce any potential bias in our record linkage (and the resulting estimates) introduced by the issues described earlier, we developed a multifaceted matching process that includes *both* address (two matching algorithms) and person-level matching (one matching algorithm). In short, we identified most HUD-assisted housing units in the ACS using address matching. We identified additional HUD-assisted housing units in the ACS by matching the ACS household roster to rosters of HUD-assisted households. When we found at least one ACS household member who was also a member of a HUD-assisted household, we considered that ACS housing unit to be HUD-assisted.

2.2 How Our Linked Should Be Used

[2] Information contained in the cited report is also confirmed by the author's extensive, but unpublished, study of address-matching HUD and property tax administrative records to both ACS and the American Housing Survey (AHS).

When linking a programmatic data source, such as HUD administrative records, to a survey source, such as the ACS, the linked data set has at least two main uses. The first main use is to create ACS-based summary statistics of characteristics of the HUD-assisted housing units or households present in the ACS sample. For instance, the linked HUD-ACS data set can be used to determine what percentage of HUD-assisted housing units have high-speed internet—a characteristic available in the ACS. For this type of statistical analysis to be feasible, it is sufficient that the linkage process correctly identifies an ACS housing unit as receiving HUD assistance.

The second main use of the data set is to conduct record-level comparisons. In a record-level comparison, a researcher would compare the ACS value for some characteristic with the HUD value for the same characteristic. For instance, a researcher could assess misreporting in the ACS rental payment variable by comparing the ACS respondent's reported rent with the HUD administrative record information on tenant rent. For this type of analysis to be feasible, the linkage process must (a) correctly identify a housing unit as receiving HUD assistance and (b) ensure the household represented in the ACS is, in fact, the same household represented in the linked HUD record. This situation involves a more stringent linkage condition than is necessary for the first main use described in the prior paragraph.

Our record linkage process is designed specifically for the first type of use only: producing summary statistics. When we identify an ACS housing unit as receiving HUD assistance through address matching, we make no further attempt to determine if the *household* occupying the ACS housing unit is the same household in the linked HUD administrative record. When we identify an ACS housing unit as receiving HUD assistance through person-level matching, we make no attempt to determine if *all* members of the ACS household roster match all members of the household roster on the linked HUD administrative record.

Our linkage approach was motivated by our knowledge of two characteristics of HUD-assisted households. First, within any given year, many HUD-assisted households move among units within the same HUD program, move between HUD programs, or move into and out of HUD programs. Second, not all individual members of HUD-assisted households may be listed in the HUD administrative records. This missing listing could happen for a variety of reasons, including movements of household members into and out of a household or HUD-assisted households not reporting all household members to HUD.[3]

These various types of movements will result in a mismatch between the ACS roster of household members and the linked HUD administrative record roster of household members. Limiting our linked data set to only records for which the ACS and HUD addresses match *and* for which the ACS and HUD household rosters match would have severely restricted our matched data set to the point at which the matched data would have been biased toward smaller and more stable households.

[3] Such tenants are often referred to as "ghost tenants," and HUD considers this fraud.

Given this approach and its stated limitation, an obvious question to ask is whether summary statistics will have some bias resulting from household member mismatch. We feel that this type mismatch should be random and unlikely to bias summary statistics results. For instance, consider a HUD-assisted housing unit in the 2017 ACS identified through address matching. Further consider that (a) household A occupied that unit in January through February, and they had high-speed internet; and (b) household B occupied the unit from March through December but did not have high-speed internet. If the 2017 ACS was administered to this housing unit in February 2017, then "share of HUD-assisted households with high-speed internet" would be biased toward the characteristics of household A, although household B occupied the unit for the majority of the year. Although this is technically a source of bias, we see no reason to believe any direction of the bias.

All this being said, it is feasible for researchers to restrict their analysis to only ACS housing units for which the ACS household roster matched the HUD household roster. Such use will restrict the universe of linked records, which may require a researcher to explore whether the restricted universe is representative of the full universe and develop bias-correction procedures.

3. Data

This section describes the two data sources being matched (ACS and HUD rental housing assistance administrative records) and two Census Bureau data sources integral to the matching process: the Master Address file and the Numident file. This section concludes by describing the chronological alignment of the HUD administrative records and the ACS.

3.1 The ACS Internal Use File Microdata

The Census Bureau produces three main products from each annual ACS: public use summary tables, public use microdata sample (PUMS) microdata, and internal use files (IUFs) microdata.

ACS IUFs differ from the PUMS microdata in at least four important ways.[4] First, ACS IUFs contain precise location information (that is, housing unit address), which can be used to match to other data sources. Secord, ACS IUFs contain respondent names, which can be used to link to other data sources. Third, ACS IUFs are a "predisclosure avoidance" version of the survey data. This characteristic means the ACS data have undergone consistency edits and include imputations for missing data but have yet to undergo the process of applying disclosure avoidance techniques, such as top-coding, bottom-coding, re-coding, rounding, and swapping. Finally, ACS IUF microdata include all the survey responses, whereas ACS PUMS microdata typically include about two-thirds of the actual number of respondents (U.S. Census Bureau, 2009).

3.2 HUD's Rental Housing Assistance Administrative Records

As discussed in Section 1, HUD oversees several rental housing assistance programs. HUD provides funding, sets overall program policy, and defines information collection requirements.

[4] ACS IUFs represent, in practice, a series of survey data files and attachable crosswalks that contain additional information about this housing unit or individual.

Among the information collection requirements for rental housing assistance programs are unit-level address and the name, date of birth, and Social Security number for each member of a HUD-assisted household.

Public housing agencies (PHAs) are responsible for the day-to-day administration of the PH and HCV programs, including information collection requirements. In the PBRA program, private property owners play a role similar to that of a PHA. HUD requires PHAs and property owners to use HUD forms 50058 (for PH and HCVP) and 50059 (for PBRA) to collect information from tenants who currently are or seek to participate in a rental assistance program. Although the data collection forms are paper, virtually all PHAs and property owners use their preferred software vendor to implement an electronic version of the forms. The electronic data are transferred to HUD on a daily basis, and HUD uses them to monitor program compliance and performance.

It is within this "distributed program administration" model where a key data quality issue emerges: poor address quality. Both HUD forms 50058 and 50059 require a PHA or property owner to input the unit-level address where the HUD-assisted household is currently residing or the address for which a new HUD-assisted household will reside. HUD does not require that the PHA or property owners check if the address is properly formatted or even valid, however. Simply put, HUD currently lacks a real-time system or service for verifying address quality.

Because it is necessary to have an accurate unit-level address for the administration of the HCVP, HCVP addresses are typically high quality and unique. Nonetheless, both PH and PBRA suffer from poor address quality. Because addresses are a common identifier used to link administrative records to other data sources, address quality directly affects linkage quality. High-quality addresses are those that are considered *deliverable* by the U.S. Postal Service and *unique* among all other addresses within a database of addresses. Obtaining high-quality addresses for HUD-assisted housing units is an especially difficult problem. Many HUD-assisted housing units are apartments within a larger building, meaning a high-quality address is typically composed of a building number, a street name, and an apartment number. Unit numbers can be problematic for address matching, as they often are incorrect or are provided in a variety of formats.

Recognizing these issues, HUD made efforts to improve the quality of their unit-level addresses for PH units, as detailed in Bucholtz and Pamnani (2016). In short, Bucholtz and Pamnani conducted extensive address cleaning to ensure that the unit-level PH addresses were deliverable and unique. HUD was unable to systematically improve the unit-level address quality for PBRA addresses.

The result is that both PH and HCVP addresses are generally high quality, but PBRA addresses lack unit-level address quality. As these quality issues affect the quality of record linkage efforts in subsequent sections, we describe a multifaceted record linkage process developed by HUD and Census Bureau staff that takes advantage of both address- and person-level identifiers.

3.3 The Census's Bureau Master Address File

The Census Bureau's Master Address File (MAF) is a collection of all addresses in the United States. The MAF was originally built for the 2000 Decennial Census (National Research Council, 2004) using addresses from the 1990 Decennial Census and the U.S. Postal Service (USPS) Delivery Sequence File (DSF). After the 2000 Decennial Census, the MAF was integrated with the Census Bureau's TIGER database. Between 2000 and 2010, the Census Bureau spent approximately $1.4 billion to improve the MAF/TIGER database (U.S. Census Bureau, Office of Inspector General, 2012). Currently, the MAF is updated twice a year using the USPS DSF, other USPS information, and information gathered during other Census survey and Decennial preparation operations (U.S. Census Bureau, 2014b). The MAF also serves as the sampling frame for most major household surveys administered by the Census Bureau, including the ACS. We leverage this feature of the ACS during our linkage process.

A key feature of the MAF is that every address in the MAF is assigned a unique key called a MAF identifier, or MAFID.

It is important to note that the MAF is not without errors, and the Census Bureau makes extensive efforts to understand where these errors may occur and target them for future correction (Heim and Raim, 2016; Pennington and Colosi, 2014). Although the reasons for MAF errors are numerous and complex, they generally fall into two categories: real housing units not included on the MAF (false negatives) or addresses on the MAF that no longer serve as housing units (false positives).

3.4 The Census's Bureau Numident File

The Social Security Administration (SSA) maintains a file for each SSN called the Numident file. This file includes the name, date, and place of birth of the holder; and the parents' names and SSNs. It also contains all transactions for an SSN, including a death.

Using the SSA's Numident file as a base, the Census Bureau builds its own version of a Numident file on a regular basis. The Census Bureau's Numident file is augmented with information from other federal and state administrative records, including current address and household composition. All transactions related to a given SSN are resolved to produce a Census Numident file containing one data record for each SSN (Wagner and Layne, 2014). Critically, the Census Bureau's Numident file contains a pseudo-code SSN called a Protected Identification Key, or PIK. In other words, the Census Bureau has assigned a PIK to all persons with an SSN.

3.5 Aligning the HUD Annual Extract to the ACS Data Collection Period

The ACS is an annual survey; however, in practice, data are collected throughout the year, with approximately one-twelfth of the annual sample enumerated each month (U.S. Bureau of the Census, 2014a). When the Census Bureau produces ACS estimates, the estimates are considered annual, meaning that they reflect the average estimate throughout the year.

There are two important caveats to the general statement that ACS estimates are annual. First, the total population and housing unit estimates are *controlled* to the corresponding estimate of total people and housing units produced by the Census Bureau Population Division. The control estimates are as of July 1 of the survey year. Second, person, family, and household income

estimates are adjusted to reflect the month in which they were collected. For instance, a household reporting an income of $100,000 in January of a survey year may have a different reported income estimate from a household reporting the same income level but in December.

The HUD rental assistance data contained in HUD's housing administrative record system are updated daily; however, HUD supplies rental assistance data from its core systems to the Census Bureau only once per year. The HUD rental assistance extract supplied to the Census Bureau is a snapshot as of approximately February 1 of a calendar year and includes any record that was active at any point from January 1 to December 31 of the preceding calendar year. For instance, the 2017 HUD rental assistance snapshot, which is produced in February 2018, includes any record that was active at any point between January 1, 2017, and December 31, 2017.

We considered two options for ensuring that our data sources chronologically align. The first option was to restrict the HUD rental assistance data to records that were active on July 1 of a calendar year. This option is motivated by the fact that ACS housing unit count estimates are controlled to total housing units, as of July 1, so HUD rental assistance data are limited to only those records that were active on that day,

The second option, and the one we chose to adopt, was to include any HUD rental assistance record that was active at any point during the calendar year. In choosing this option, we recognized that the most likely use of the matched information from HUD is to produce share estimates, such as "percentage of HUD-assisted housing units with a veteran." In this sense, using all HUD rental assistance records that were active at any time during the year means that share estimates reflect a program annual average.

4. Record Linkage Process

This section provides a step-by-step description of the four steps in the record linkage process used to match HUD-assisted housing units and households to the ACS housing units. First, the HUD data were cleaned. In the second step, HUD-assisted housing units were address matched to ACS housing units. In the third step, HUD-assisted housing units were matched to ACS housing units using person matching. In the fourth step, we made a final linkage determination based on the results of steps 2 and 3.

4.1 Step 1: Cleaning the HUD Administrative Records

Record linkage requires extensive data pre-processing (Playford et al., 2016). Before being matched to the MAF, the Census Bureau uses standardization software that edits and standardizes the HUD administrative data unit addresses.

4.2 Step 2a: Probabilistically Linking PH and HCVP Administrative Records to the MAF (Linkage Algorithm #1)

In this step, an attempt is made to link all public housing and Housing Choice Voucher program records to a Master Address file record, thereby attaching a unique match key to each HUD record. As explained in Section 3.2, these records have high enough address quality data to allow for this. The Census Bureau performs linking to the MAF in a three-step process. The first step

(or pass) attempts to find a string match using the full address (building number, street name, and apartment number). The second step attempts to find a string match using the rural route address. The third step attempts to find a string match using the basic street address (BSA), which is composed only of the building number and street name. To reduce the computational load, each step has its own blocking procedures, ranging from building number and ZIP code to just ZIP code (Wagner and Layne, 2014).

It is important to note that the string comparator algorithms produce a match weight (often referred to as a match score) that quantifies the likelihood an individual MAF address is a match to an address from an individual HUD administrative record. For example, the string comparator algorithm might output that two addresses are a 95-percent-likely match. Census Bureau experts determine the threshold (that is, minimum match weight) that constitutes a match. The MAF record with the highest match weight that is above the minimum match weight threshold is considered the linked record. Each PH and HCVP administrative record that links to a MAF record then receives the MAF identification number (MAFID) representing the linked MAF record.

Unfortunately, many PH and HCV program administrative records have addresses that either (1) cannot be linked to any address in the MAF (match weight equals zero); (2) can be linked, but the match weight does not meet the minimum threshold to be considered a link; or (3) share a link to the same MAFID. HUD administrative records without a MAFID cannot be matched to the ACS using the process described above because a link to the MAF is a necessary condition for that linkage. For HUD administrative records with duplicate MAFIDs, only one record is kept. Duplicated MAFIDs make up a very small portion of PH and HCVP records. By keeping one record of the duplicated MAFID, we assume that duplicated MAFIDs have at least one correct link. Records without a unique MAFID are still eligible to be matched to the ACS using the process described in Section 4.7.

Table 4.1 describes the results of linking PH and HCVP administrative records to the MAF. About 90 percent of PH and HCVP records were able to uniquely match to the MAF.

Table 4.1 Share of PH and HCVP Records Linking to the MAF

	2011	2012	2013	2014	2015	2016	2017
Number of PH or HCVP records	3,257,000	3,245,000	3,238,000	3,240,000	3,263,000	3,314,000	3,290,000
Share of records linked to a unique MAF record	90.3%	91.1%	91.0%	91.1%	91.1%	90.6%	91.0%
Share of records NOT linked to a unique MAF record	9.7%	8.9%	9.0%	8.9%	8.9%	9.4%	9.0%

HCVP = Housing Choice Voucher program. MAF = Master Address file. PH = public housing.

4.3 Step 2b: Probabilistically Linking PBRA Administrative Records to the MAF (Linkage Algorithm #2)

Because of the way the PBRA program is administered, many PBRA unit-level addresses are not high quality (that is, deliverable or unique.) There are many reasons for this, but often the reason is that some PBRA unit-level addresses are missing unique apartment numbers. Without a unique apartment number, PBRA units cannot be linked to a unique MAF record.

Importantly, the PBRA program is a contract-based program, meaning that HUD enters into a contract with the owner of a building to provide housing units. It is very often the case that all or nearly all the units in the building are under contract. For instance, an owner of a 50-unit building is likely to enter into a PBRA contract with HUD to provide 50 units.

To overcome poor address quality for PBRA administrative records, we developed a technique to impute missing apartment numbers. In a nutshell, we first attempted to link all PBRA units to the MAF based on their full address, if present. For PBRA units that could not be linked to the MAF, we then exploited the fact that PBRA addresses, although often missing apartment numbers, almost always have a BSA. We imputed missing apartment numbers using addresses from that *same* basic street address found in the MAF.

For instance, if the PBRA data shows 20 units at 123 Main Street that are missing apartment numbers, we extract all the full MAF addresses (that is, basic street address and apartment number) at 123 Main Street, then randomly assign each of the 20 units to one of the full MAF addresses at 123 Main Street, ensuring that no two PBRA addresses are assigned to the same full MAF address.

The result of this process is that most PBRA units are linked to the MAF and receive the MAF identification number (MAFID) representing the linked MAF record. Of course, some PBRA units cannot be linked to the MAF, for the same reasons described in Section 3.3.

Table 4.2 shows the share of PBRA records that had their unit-level address imputed. In most years, less than one-half of the PBRA addresses can be uniquely linked to the MAF "as is." Between 22 and 34 percent were linked to the MAF only after imputing an apartment number. Between 22 and 24 percent could not be uniquely linked to the MAF.

Table 4.2 Share of PBRA Records Linking to the MAF

	2011	2012	2013	2014	2015	2016	2017
Number of PBRA records	1,511,000	1,438,000	1,477,000	1,458,000	1,494,000	1,446,000	1,453,000
Share of records linked to a unique MAF record without imputation	43.4%	44.0%	43.4%	45.0%	43.2%	52.4%	53.1%

Share of records linked to a unique MAF record using imputation	33.4%	34.3%	34.1%	32.4%	32.8%	26.8%	26.5%
Share of records NOT linked to a unique MAF record	23.2%	21.7%	22.5%	22.6%	24.0%	20.8%	20.4%

MAF = Master Address file. PBRA = project-based rental assistance.

4.4 Step 2c: Deterministically linking HUD administrative records to the ACS using MAFID

Once all records have a unique MAFID, they are deterministically linked to ACS housing units. This determination is feasible because of a key feature of the ACS previously mentioned in Section 3.3: the MAF serves as the sample frame for the ACS. That fact means that each ACS housing unit is "drawn" from the MAF and hence comes with a corresponding MAFID. Deterministically linking the HUD administrative records and the ACS is carried out via the MAFID match key.

4.5 Step 3a: Deterministically linking person-level HUD administrative records to the Numident file

For each HUD-assisted housing unit, HUD supplied a person-level roster to the Census Bureau, which includes Social Security numbers. The Census Bureau then linked the HUD person-level rosters to their Numident file using SSN as the match key. The result of this link is that each person in a HUD household receives a unique Protected Identification Key from the Numident file.

4.6 Step 3b: Probabilistically linking person-level ACS records to the Numident file (Linkage Algorithm #3)

On the ACS questionnaire, respondents are asked to provide their name, date of birth, and gender; however, they are not asked to provide their SSN.

The Census Bureau has developed a process to link survey respondents to the Census Bureau's Numident file using information provided by the survey respondent. This process is embedded in a system called the Person Identification Validation System, or PVS. As discussed in great detail by Wagner and Layne (2014), the PVS uses the respondent's name, gender, address, date of birth, and household relationship to probabilistically link a respondent to the Numident file, thereby assigning (if possible) each ACS respondent a PIK.

It should be noted that the PVS includes thresholds on what level match score is considered a match. Just as with the MAF linking processes described in Section 4.3, the PVS thresholds are set by Census experts and have undergone extensive internal review. It is beyond the scope of

this project to perform sensitivity analysis on these thresholds. For an evaluation of PVS matching applied to ACS respondents, see Mulrow et al. (2011).

Table 4.3 shows the share of ACS records that have been linked to the Numident file. About 94 percent of ACS person-level records have been linked to the Numident, and about 96 percent of ACS households have at least one household member that has been linked to Numident.

Table 4.3 Share of ACS Person-level Records Linking to the Numident

	2011	2012	2013	2014	2015	2016	2017
Number of ACS person-level records*	4,879,000	5,446,000	5,105,000	5,325,044	5,270,000	5,082,000	4,881,000
Share of ACS person-level records linked to the Numident	93.8%	93.8%	94.3%	94.4%	94.4%	93.8%	93.7%
Share of ACS households with at least one person-level record linked to the Numident	95.3%	95.3%	95.9%	96.0%	96.1%	96.1%	95.9%

* Individuals in Puerto Rico have been excluded.
ACS = American Community Survey.

4.7 Step 3c: Deterministically Linking Person-level HUD Administrative Records to the ACS

Once HUD-assisted household members are assigned PIKs (Section 4.5) and ACS respondents are assigned PIKs (Section 4.6), they can be linked using PIK as the match key. We exclude any HUD-assisted household member that was listed in a household previously linked by MAFID. Once the records were linked by PIK, we removed links for which the county in the HUD administrative data did not match the county in the ACS data.

The remaining issue is to decide how many linked household members are required to declare a household-level match. To understand this decision, consider a three-person HUD-assisted household composed of a woman and her two children. Further suppose that this household responds to the ACS. If HUD's administrative records were complete, the household fully responded accurately to the ACS, the Census Bureau's Numident file was complete, and the PVS process functioned as intended, then each of the three HUD-assisted household members should link to their respective ACS record.

Of course, that is a lot of "ifs." After consideration, we decided that a link between at least one member of the HUD-assisted household roster and at least one member of the ACS household roster was sufficient to identify the ACS housing unit as HUD-assisted. The effect of this

decision is expected to be minimal, as just under 50 percent of all HUD-assisted households are single-person households.[5]

4.8 Step 4: Final Linkage Determination

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, as long as the HUD administrative record and the ACS record are in the same county.

4.9 Linkage Rates by Type of Match

Table 4.4 shows the breakdown of cases, by year, that were MAF-matched versus those that were PIK-matched. As address data improved in following 2015, the rate of cases MAF-matched has increased.

Table 4.4 Breakdown of ACS Records Linked to HUD Record, by Match Type

	2011	2012	2013	2014	2015	2016	2017
ACS records linked by MAF-match	60,000 81.6%	63,500 81.9%	55,500 82.2%	58,000 82.9%	57,000 82.6%	54,500 83.8%	51,500 84.4%
ACS records linked by PIK-match	13,500 18.4%	14,000 18.1%	12,000 17.8%	12,000 17.1%	12,000 17.4%	10,500 16.2%	9,500 15.6%
Total	73,500	77,500	67,500	70,000	69,000	65,000	61,000

ACS = American Community Survey. MAF = Master Address file. PIK = Personal Identification Key.

5. Record Linkage Quality Discussion

In this section, we briefly discuss the potential for and nature of linkage errors in the *probabilistic* steps of our multifaceted matching approach.

5.1 Sources of Linkage Failure When Probabilistically Linking PH and HCVP Administrative Records to the MAF

Linking PH and HCVP administrative records to the MAF can fail for a few reasons. First, HUD's unit-level addresses may not be of sufficient quality to be uniquely linked to a MAF address. Second, MAF addresses may not be of sufficient quality to be uniquely linked to a HUD unit-level address. Third, the linking algorithm may fail to designate two addresses as linked (false negative), or it may incorrectly designate two addresses as linked when in fact they do not represent the same housing unit (false positive).

Table 4.1 revealed that about 9 percent of PH and HCVP records could not be uniquely linked to the MAF. That means that, without any further linking effort beyond address linking, only about

[5] Based on the authors' calculation of HUD administrative records for 2017.

91 percent of the ACS records that were indeed in a PH or HCV program would be identified as such. Therefore, the expected false negative rate for PH and HCV program records is 9 percent.

To calculate a false positive rate, we drew a simple random sample of 250 PH and HCV program unit-level addresses, and then we manually reviewed them to determine if the HUD unit-level address and the MAF address were the same. In this "gold standard" data set, 249 of the 250 pairs of addresses were determined to be the same. This errant link was for a PH unit that was linked to a MAF record of another unit in the same BSA. All units within a public housing complex fall under a PH program. As such, the false positive rate from this part of the overall linking process is negligible.

5.2 Sources of Linkage Failure When Probabilistically Linking PBRA Administrative Records to the MAF

PBRA records may fail to link to MAF records or may incorrectly link to MAF records for the same reasons mentioned in Section 5.1. In fact, Table 4.2 showed that less than one-half of the PBRA records could be linked to the MAF, so without further efforts to link PBRA records to the MAF, less than 50 percent of the ACS records that are in PBRA would be identified as such; therefore, the expected false negative rate is at least 50 percent.

To calculate a false positive rate, the same protocol from Section 5.1 was used. A simple random sample of 250 PBRA addresses that linked to the MAF were drawn. Then, the PBRA address was manually compared with the MAF address. In this PBRA "gold standard" dataset, 247 of the 250 pairs of PBRA addresses were determined to be the same. In all three cases, the link was for a different unit number in the same BSA. An imputation technique similar to that used for other PBRA cases described in Section 4.3 was applied. As such, the false positive rate from this part of the overall linking process is negligible.

For 20 to 30 percent of the PBRA records that were linked to the MAF using the special address imputation technique described in Section 4.3, false positives and false negatives are expected due to the nature of the imputation. Because the PBRA records lack valid unit numbers, it is not feasible to directly calculate false positive or false negative rates.

Despite the inability to *directly* calculate false positive or negative rates, these rates can be estimated. To understand how, consider the scenario in which a 10-unit PBRA building is located at 123 Main Street, but the PBRA records lack apartment numbers. To impute apartment numbers, all the unit-level records at 123 Main Street are extracted from the MAF. If there are exactly 10 MAF records at 123 Main Street, then the imputation of apartment numbers is in fact exact. If one of these apartments is in the ACS sample, a PBRA record will link to it. In this case, the false negative and false positive rates will be zero.

By contrast, if there are 20 MAF records at 123 Main Street, then the imputation process randomly selects only 10 of the 20 MAF records to link to the PBRA records. Due to the random selection, on average, one-half of the links will be correct, one-fourth will be false positives, and one-fourth will be false negatives. Although false negative and false positive links will occur for

the imputed links, the 10 PBRA housing units at 123 Main Street (and the households that occupy them) are expected to be substantially similar to the 10 non-PBRA units.

Using all the PBRA records that were linked to the MAF using imputation, it is estimated that 18 percent are false positives and 18 percent are false negatives. These low rates, combined with the description provided in the prior paragraph, suggest that false negatives and false positives "wash out" on net.

5.3 Sources of Linkage Failure When Probabilistically Linking Person-level ACS Records to the Numident File

Linking ACS person-level records to the Numident can fail for a few reasons. First, an ACS respondent may provide incorrect or incomplete information for one or more of the characteristics used by the PVS, thereby leading to either a failure to link to a record in the Numident file or a link to an incorrect record in the Numident file. Second, the Numident file may have incorrect information. Third, the PVS algorithms may fail to link to records or may incorrectly link to records.

Table 4.3 shows that about 6 percent of ACS person-level records cannot be linked to the Numident file; however, the statistic that is important for this linkage process is the share of ACS households that do not have at least one person-level record that can be linked to the Numident file. That share is about 5 percent, and it represents a false negative rate.

The PVS can incorrectly link an ACS record to the Numident file, which could create a false positive. This result is caused by the presence of missing demographic information in the ACS, such as name, date of birth, or address. It is difficult to assess the false positive rate of linking ACS records to the Numident file, as no truth deck exists. Previous work in using verified survey data as a truth deck found the false positive rate to be rare (Mulrow et al., 2011).

6. Record Linkage Quality Assessment

In this section, we present linkage quality assessment metrics by way of global totals agreement, discuss limitations when making record-level comparisons between ACS housing unit or household characteristics and HUD housing unit or household characteristics, and discuss a method to overcome any potential bias in the final estimates.

6.1 Global Totals Agreement

To determine whether the multifaceted linking process performed well, we compared the pre-linking known HUD-assisted record count with the post-linking ACS weighted estimate of HUD-assisted housing units. For instance, suppose HUD provides the Census Bureau with approximately 4,715,000 unit-level records for occupied or vacant-but-available-for-occupancy units. If the linking algorithms perform well, the post-link ACS weighted estimate of HUD-assisted units should equal the pre-link known record count.

Table 6.1 presents linking quality metrics for 2011 through 2017. For example, the table shows that HUD provided the Census Bureau with 4.72 million HUD-assisted housing unit records in 2013. When these records were linked to ACS housing units, the weighted estimate of HUD-

assisted housing units was 4.70 million, or 99.8 percent of the real total. Across all years, the ACS-weighted estimate as a share of HUD-assisted housing units ranges from 99.8 percent to 101.1 percent. This result suggests that the linking process identifies nearly all HUD-assisted housing units within the ACS.

The difference between the ACS-weighted estimate of HUD-assisted housing units and the actual total (100 percent) can be interpreted as an overall false negative (seen when the share is less than 100 percent) or false positive rate (seen when the share is greater than 100 percent). The false negative/positive rate overall ranges from 0.6 to 1.6 percent.

The table below also presents the linking quality metrics by HUD program. The results show that the linking process for both PH and HCV program performs well, whereas the linking process for PBRA performs slightly worse but still well enough to consider the process sound. The ACS-weighted estimate of public housing units ranges from 97.7 percent to 102.3 percent of the true total. For the HCV program, the estimate ranges from 97.7 to 100.9 percent. For PBRA, the estimate ranges from 92.0 to 95.2 percent.

Finally, we must note that because the ACS is a sample-based survey, all estimates are subject to sampling error, including estimates of HUD-assisted housing units. The 90-percent margin of error for PBRA never includes 100 percent, which means we have false negatives. Estimates for all program types improved across the years.

Table 6.1 Results of ACS/HUD Administrative Matching

		All	PH	HCV	PBRA
2011	HUD records provided to Census	4,768,000	1,054,000	2,202,000	1,511,000
	ACS estimate of HUD-assisted households	4,624,000	1,030,000	2,204,000	1,390,000
	ACS 90-percent margin of error	0.7%	1.6%	1.1%	1.3%
	ACS estimate as share of HUD records	97.0%	97.7%	100.1%	92.0%
2012	HUD records provided to Census	4,683,000	1,035,000	2,210,000	1,438,000
	ACS estimate of HUD-assisted households	4,653,000	1,054,000	2,230,000	1,369,000
	ACS 90-percent margin of error	0.7%	1.4%	0.9%	1.3%
	ACS estimate as share of HUD records	99.4%	101.8%	100.9%	95.2%
2013	HUD records provided to Census	4,715,000	1,044,000	2,194,000	1,477,000
	ACS estimate of HUD-assisted households	4,578,000	1,034,000	2,156,000	1,388,000
	ACS 90-percent margin of error	0.7%	1.3%	1.1%	1.4%
	ACS estimate as share of HUD records	97.1%	99.0%	98.3%	94.0%
2014	HUD records provided to Census	4,699,000	1,052,000	2,188,000	1,459,000
	ACS estimate of HUD-assisted households	4,612,000	1,069,000	2,154,000	1,389,000
	ACS 90-percent margin of error	0.7%	1.4%	1.1%	1.1%
	ACS estimate as share of HUD records	98.1%	101.6%	98.4%	95.2%

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 90-percent margin of error	0.7%	1.4%	1.2%	1.3%
	ACS estimate as share of HUD records	98.3%	102.3%	99.6%	93.7%
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 90-percent margin of error	0.7%	1.5%	1.1%	1.1%
	ACS estimate as share of HUD records	97.1%	98.7%	97.7%	95.0%
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 90-percent margin of error	0.8%	1.5%	1.1%	1.3%
	ACS estimate as share of HUD records	97.3%	100.3%	98.1%	94.1%

ACS = American Community Survey. HCV = housing choice voucher. PBRA = project-based rental assistance. PH = public housing.

Given those results, a reasonable conclusion is that the matching algorithm performed well enough to ensure that the ACS housing unit cases flagged as HUD-assisted units are a representative cross-section of all possible ACS cases that are HUD-assisted units. In statistical terms, although there are false negatives, their omission will not necessarily result in biased estimates of *housing or household characteristic*s of HUD-assisted households.

6.2 Correcting for Remaining Bias

Although we concluded Section 6.1 by indicating our belief that the linked data produces estimates with little bias, we acknowledge the possibility of some bias in the linkage process that could result in biased estimates. To the extent that any bias in estimates is produced from the linked data, we believe that a simple method to correct bias is to control estimates to official HUD control totals. Our suggested method is to control linked data estimates to official HUD totals by the combination of HUD program (3), Census Division (9), and presence/absence of an elderly household member (2). For instance, for 2015, a control totals table resembles Table 6.2. These control total tables are available to researchers accessing the data.

Table 6.2 HUD Control Totals for 2017

Census Division	Public Housing Elderly	Public Housing Non-Elderly	HCV Elderly	HCV Non-Elderly	PBRA Elderly	PBRA Non-Elderly
New England	29,586	36,155	38,964	117,360	77,813	50,967
Middle Atlantic	115,962	173,930	110,541	259,903	149,816	98,909

East North Central	40,370	88,256	58,941	241,856	135,673	145,740
West North Central	21,149	37,077	25,743	98,499	52,286	53,803
South Atlantic	39,966	104,523	83,650	274,170	113,227	123,409
East South Central	23,746	68,310	19,528	103,529	41,457	59,536
West South Central	25,700	63,436	49,276	186,547	47,534	62,413
Mountain	7,928	17,268	25,881	76,836	30,519	28,053
Pacific	17,170	33,813	141,321	254,444	104,921	52,049

HCV = housing choice voucher. PBRA = project-based rental assistance.

6.3 Caution When Performing Record-level Comparisons

In Section 2.2, we discussed a limitation on our linked data. Our linkage process was designed only to identify ACS housing units as receiving HUD assistance. That process means that the linked data set can be used to produce estimates of ACS characteristics for HUD-assisted housing units. We made no attempt to determine if the ACS household and the HUD household contained the same members, however. That fact means that our linked data set is not necessarily suitable for comparing the values of housing unit or household characteristics that are common to the ACS and HUD administrative data, such as rent payment or demographics.

We acknowledge that bias may exist in our linked data set due to the linkage process. We further acknowledge that one way to measure bias is to review the distributions of characteristics in the linked data that are common to both data sets. For instance, both data sets include "number of bedrooms," and this characteristic is likely to be unaffected in instances in which the ACS roster of household members does not match the HUD roster of household members; however, we elected not to compare these common attributes. The ACS data are respondent-reported and contain errors. For example, we know that all individuals in public housing are renters; however, we see ACS respondents in public housing report that they own their housing unit. It was our feeling that differences in distributions of common characteristics could be entirely driven by respondent error, and there was no way to separate actual bias from respondent error.

7. Conclusion

We developed a multifaceted approach to identify HUD-assisted rental housing units in the ACS. To do this, we linked administrative data from HUD's rental housing assistance programs to ACS housing units using address matching and person-level matching for years 2011 through 2017. In each year of the ACS, we were able to identify between 61,000 and 78,000 ACS housing units as being HUD-assisted.

By linking the two data sources, we can learn more about HUD-assisted households without having to conduct an expensive, one-off survey. For instance, we could determine the estimated number of veterans in HUD-assisted households or the estimated percentage of HUD-assisted households without broadband internet.

These data can be used in many possible ways. Our goal with this project was to develop the matching process and build out the matched data sets so researchers can further explore the data. Access to the ACS/HUD internal use files, or IUFs, are available to researchers through a Census Bureau Research Data Center (RDC) who obtain Special Sworn Status and approval for their project. The process is as follows:

1. Identify the RDC nearest to you (https://www.census.gov/fsrdc).
2. Contact the RDC administrator to explain your interest in using the linked ACS/HUD IUF.
3. Complete a proposal following the proposal guidelines (https://www.census.gov/content/dam/Census/programs-surveys/center-for-economic-studies/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 RDC.
6. Submit your results to the RDC for clearance.

ACS/HUD IUFs users should note that all estimates produced from the ACS/HUD IUFs must go through Census Bureau disclosure review.

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.

Brucker, Debra L., Veronica Helms, and Teresa Souza. 2017. "Health and Health Services Access Among Adults with Disabilities Who Receive Federal Housing Assistance," *Housing Policy Debate* 28 (2): 248–266.

Brummet, Quentin, Denise Flanagan-Dyle, Joshua. Mitchell, John Voorheis, Laura Erhard, and Brett McBride. 2018. *Investigating the Use of Administrative Records in the Consumer Expenditure Survey*. Washington, DC: U.S. Census Bureau, Center for Administrative Records Research and Applications.

Bucholtz, Shawn J., and Pramod. Pamnani. 2016. *Constructing Unit-Level Addresses in the Public and Indian Housing Information Center.* AHS Redesign Paper #4. Washington, DC: 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 & Health* 2 (2): 111–119.

Chetty, Raj, Nathanial Hendren, and Lawrence Katz. 2015. *The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment.* NBER Working Paper No. 21156. Cambridge, MA: National Bureau of Economic Research.

Clark, Sandra L., R.C. Sawyer, Amanda Klimek, Christopher Mazur, William Chapin, and Ellen Wilson. 2018. *Housing Administrative Record Simulation.* Suitland, MD: U.S. Census Bureau.

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

Heim, Krista, and Andrew M. Raim. 2016. "Predicting Coverage Error on the Master Address File Using Spatial Modeling Methods at the Block Level," *Joint Statistical Meetings, Survey Research Methods Section*: 1541–1555.

Jarmin, Ron C., C.J. Krizan, and Adela Luque. 2016. *Small Business Growth and Failure During the Great Recession: The Role of House Prices, Race and Gender.* No. 2016-08. Washington, DC: Center for Economic Studies, U.S. Census Bureau.

Luque, Adela, and Renuka Bhaskar. 2014. *2010 American Community Survey Match Study.* Center for Administrative Records Research and Applications Working Paper CARRA-WP-2014-03. Washington, DC: U.S. Census Bureau.

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): 1-8.

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.

Mulrow, Edward, Ali Mushtaq, Santanu Pramanik, and Angela Fontes. 2011. *Assessment of the US Census Bureau's Person Identification Validation System.* Chicago, IL: NORC at the University of Chicago.

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

O'Hara, Amy B., 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.

Pennington, Robin A., and Robert. Colosi. 2014. *Targeted Address Canvassing and Master Address File Modeling and Testing.* Suitland, MD: U.S. Census Bureau.

Playford, Christopher J., Vernon Gayle, Roxanne Connelly, and Alasdair J. Gray. 2016. "Administrative Social Science Data: The Challenge of Reproducible Research," *Big Data & Society* 3 (2): 1-13.

Rastogi, Sonya, Amy B. O'Hara, James Noon, Ellen A. Zapata, Cindy Espinoza, Leah B. Marshall, Teresa A. Schellhamer, and J. David Brown. 2012. *2010 Census Match Study.* Washington, DC: U.S. Census Bureau, Center for Administrative Records Research and Applications Report.

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

U.S. Census Bureau. 2009. *A Compass for Understanding and Using American Community Survey Data.* Suitland, MD: U.S. Census Bureau.

———. 2014a. *American Community Survey Design and Methodology (January 2014). Chapter 7: Data Collection and Capture for Housing Units.* Washington, DC: U.S. Census Bureau.

———. 2014b. *American Community Survey Design and Methodology Report (January 2014). Chapter 3: Frame Development.* Suitland, MD: U.S. Census Bureau. https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch03_2014.pdf.

U.S. Census Bureau, Office of Inspector General. 2012. *High-Quality Maps and Accurate Addresses Are Needed to Achieve Census 2020 Cost-Saving Goals.* Suitland, MD: U.S. Census Bureau.

U.S. Commission on Evidence-based Policymaking (CEP). 2017. *The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking.* Washington, DC: U.S. Commission on Evidence-based Policymaking.

Voorheis, John. 2017. *Air Quality, Human Capital Formation and the Long-term Effects of Environmental Inequality at Birth.* CARRA Working Paper No. 2017-05. Washington, DC: U.S. Census Bureau, Center for Economic Studies.

Wagner, D., and M. Layne. 2014. *The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications'(CARRA) Record Linkage Software.* Working Paper No. CARRA WP 2014-01. Washington, DC: U.S. Census Bureau, Center for Economic Studies.

Wang, Teresa W., Veronica E. Helms, Peter J. Ashley, and Brian A. King. 2015. "Characteristics and Correlates of Cigarette Smoking Status Among US Adults Receiving Federal Housing Assistance," *Preventing Chronic Disease* 15: E48.

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.

U.S. Department of Housing and Urban Development
Office of Policy Development and Research
Washington, DC 20410-6000

August 2019