UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

February 8, 2019

2018 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES # ACS18-RER-03

CENTER FOR ECONOMIC STUDIES WORKING PAPER # 18-52

MEMORANDUM FOR Jennifer Ortman
Assistant Division Chief for Survey Methods and Measures

From: Sonya Porter
Acting Assistant Center Chief for Demographic Research, Center for Economic Studies

Prepared by: Michaela Dillon, Kevin Rinz and John Voorheis
Center for Economic Studies

Subject: Use of Administrative Records to Replace or Enhance the Facilities & Fuel Type Questions on the American Community Survey

Attached is the Center for Economic Studies (CES) Research and Evaluation report, "Use of Administrative Records to Replace or Enhance the Facilities & Fuel Type Questions on the American Community Survey." We conducted this evaluation to assess the potential for using commercial vendor data sourced from multiple listing service (MLS) records to replace or supplement the facilities or heating fuel type questions on the American Community Survey (ACS). If you have any questions about this report, please contact Sonya Porter at 301-763-6038, Michaela Dillon at 301-763-3567, Kevin Rinz at 301-763-6735, or John Voorheis at 301-763-5326.

Attachment

cc:
David Raglin (ACSO)
Nikolas Pharris-Ciurej (CES)
Edward Castro (DSSD)
Asaph Young Chun (DSSD)
Nathan Walters (SEHSD)
Ellen Wilson (SEHSD

This page intentionally left blank.

FEBRUARY 8, 2019

Use of Administrative Records to Replace or Enhance the Facilities & Fuel Type Questions on the American Community Survey

Michaela Dillon, Kevin Rinz, & John Voorheis
Center for Economic Studies

United States Census Bureau

This page intentionally left blank.

TABLE OF CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

As detailed in the report, "Agility in Action: A Snapshot of Enhancements to the American Community Survey" (U.S. Census Bureau, 2015), the Census Bureau is investigating ways to reduce the difficulty and length of the American Community Survey (ACS) using administrative records to address the burden survey participation places on respondents. The ACS questions under investigation for this report include facilities and fuel type. This research begins the evaluation of whether an identified administrative records source has data of sufficient coverage and quality to allow the removal of the facilities and fuel type question on the ACS. Alternatively, we may find the administrative record source sufficient only to serve as a supplement to data provided by respondents to fill in missing responses or enhance editing routines. A Census Bureau team will use this report and conduct additional research to make recommendations on whether each question is a good candidate for removal by using the vendor data source in its place.

This report explores the linkage of multiple listing service (MLS) data compiled by a third-party commercial vendor in 2014 to the ACS sample collected in the 2014 sample year. We use string data processing on the text in the commercial vendor's MLS file to generate measures of facilities and fuel type. We then use the linked data file to assess the presence of comparable MLS data, and the agreement of this vendor data with ACS self-reported and edited responses.

After linking the 2014 ACS to the MLS vendor data by Master Address File Identifiers (MAFIDs), we assess the coverage of the matched data, finding very low match rates. We find 13.9 percent of ACS households can be matched to the MLS vendor data in counties where the MLS vendor data has coverage. Across ACS topics, the rates at which ACS records match to non-missing information in the MLS vendor data are similar to overall match rates, as less than one percent of matched MLS records lack facilities or fuel type information. ACS households that are owner-occupied or are located in metropolitan areas are relatively more likely to match to MLS records, while households that have householders who are either very young or very old are less likely to match. However, no geographic or demographic subgroup has match rates above 30 percent.

Although the match rates are quite low, matched records do exhibit a high degree of agreement between MLS and ACS responses. When ACS records match to non-missing vendor data, the rate of agreement in linked values for facilities questions is 95.3 percent. However, the MLS data correctly identify very few cases where households lack a particular facility type. The vendor data cannot identify a single primary fuel type. However, the fuel type listed by the ACS respondent is among the multiple fuel types in the vendor data 81.3 percent of the time.

Above and beyond the low match rates, there are some additional limitations of linked data, which affect its suitability for ACS item replacement or imputation. Due to differences in the objectives and methods of collecting the data, it is difficult for vendor data to align with the ACS questions conceptually. In addition, MLS records often focus on housing units for sale as opposed to the entire universe of housing units, and cover very few rental units. Additionally, MLS records are only available for particular housing markets at this point in time, which limits the availability of the data across the entire ACS sample. These differences raise the possibility

that characteristics that vary by housing unit may not align with the same information collected by MLS records. This is of particular concern among renters in multi-family buildings.

1. BACKGROUND

Stemming from concerns about the burden that American Community Survey (ACS) participation places on respondents, the Census Bureau is looking for ways to reduce the difficulty and length of the survey with administrative records. We have identified sources of both federal and commercial data that may potentially alleviate the need to ask certain questions altogether or for a subset of the ACS sample. Work is underway to acquire new sources and assess the quality of the matching and coverage of these sources. Data from other agencies are under review to potentially replace ACS content, including the Internal Revenue Service to provide income information and the Social Security Administration for pension and disability information. The American Community Survey Office (ACSO) is consulting with stakeholders, including Congress, regarding the appropriateness of direct substitution.

Recently, the ACSO contracted with the National Opinion Research Center (NORC) to report on the availability of data sources, as well as the potential issues with those sources, as candidates for replacing/supplementing data currently collected by the ACS. Using this report (Ruggles, 2015) as well as their experience, the Center for Economic Studies (CES) identified several topics for further study based on the availability of data and likelihood of successful matching and analysis. These topics include:

- Year built
- Part of Condominium
- Tenure
- Property value
- Real estate taxes
- Have mortgage/mortgage amount
- Second mortgage/HELOC and payment
- Income in the past 12 months
- Residence one year ago
- Number of rooms/bedrooms
- Facilities
- Fuel type
- Acreage

For each topic, CES will acquire and match the administrative records to survey data, provide a report or memo describing the quality and coverage of the data source, and compare the administrative record value to ACS self-reported and edited responses. CES will document the linked file and put the research extract in the Data Management System (DMS) for future research.

This research is intended to be a first look at the various topics to document the coverage, quality, and availability of external data sources for potential ACS integration. This research will enable ACS to evaluate the potential of the replacement data sources, identify challenges, and provide direction for further research. It is an exploratory investigation of the feasibility of replacing ACS data with administrative records.

Next, the ACSO will create teams for each ACS topic identified as a potential candidate for records usage based on the results from the first phase of research. Each team will include statistical researchers, subject matter experts, and data processors that together can identify and research issues related to records usage.

The teams will make recommendations on whether each question is a good candidate for removal with the use of external data sources in its place. This recommendation will be based on an assessment of the implications of implementing such a change, considering data quality, reliability, alignment of reference periods, break in series, and the limitations of the data source affecting the suitability for use. The team will document and evaluate various options for integrating the records. For instance, for some topics, records may be better suited in assisting with imputation whereas for other topics the records may be used for direct substitution of a survey question (for all or a subset of the ACS respondent pool).

Moreover, the ACSO will gauge reactions to the intended use of external data sources from data users, stakeholders, and the public. ACSO will review current ACS mail materials to ensure proper transparency, as well as publicly share our vision in public forums such as the ACS Data Users Conference, meetings of the Association of Public Data Users (APDU), the Population Association of America (PAA), the Joint Statistical Meetings (JSM), the American Association of Public Opinion Research (AAPOR), and other public venues.

2. LITERATURE REVIEW

The report, "Review of Administrative Data Sources Relevant to the American Community Survey (Ruggles, 2015)," provides a review of data sources that could be used to replace or improve specific questions on the ACS. Its purpose is to support the work of the ACS Content Review (Chappell and Obenski 2014) by providing additional input on potential data sources that might be used to strengthen the survey, improve its content, or reduce the burden associated with its collection. While several studies emphasize the coverage of administrative records for persons, there is a subset of the literature focusing on the availability and quality of data suitable for household-level questions and surveys. For example, the 2010 American Community Survey Match Study (Luque and Bhaskar 2014) assesses the coverage of person and address administrative records data from twenty federal and third-party data sources, finding that records provide substantial coverage for persons and addresses in the 2010 ACS (92.1 and 92.7 percent, respectively).

Brummet (2014) assessed the match rate of a commercial dataset sourced from county and municipal property tax records to the 2009 American Housing Survey (AHS) and found lower coverage rates, highlighting unique patterns in housing data that any successful integration of third-party sources should consider. In particular, the match rate tends to vary by structure type and subsequently tenure. Commercial data were matched to 79.0 percent of AHS single-family housing units and 14.8 percent of multi-unit housing structures. The large difference in match rates highlights constraints on the ability of property tax records to describe significant portions of the country's households, most likely due to misalignment in the objectives for collecting housing data. For the purposes of taxation, local governments often record data at aggregate levels, capturing the entire structure and/or parcel of land rather than the housing units within. Availability of more detailed information on individual housing units depends on the mandate of state and local laws. Timing is also a documented issue when linking to tax records which may suffer from inaccuracies due to differences in the frequency of taxation across locations, and lags in the data collection process of recent construction for example.

Using the same commercial data as Brummet (2014), Seeskin (2016) adjusted for the previously noted issues by focusing on single-family, owner-occupied homes and found 69.1 percent of households in the 2010 ACS match to a commercial data record with non-missing information. Seeskin showed the distribution of property tax values is similar to those reported or edited in the ACS, and that slightly greater variation occurs in the tails. Additionally, the linked data has a Pearson correlation of 0.724, indicating a strong, positive linear relationship between the linked property tax values. While property tax data is not perfect, studies such as Zanutto and Zaslavsky (2002) support the use of tax records for imputation and nonresponse adjustments. According to Manski (2007), use of this data in models of nonresponse can be further improved by relaxing the assumption that missing data is randomly distributed.

There is evidence that the quality of the matched data varies systematically along several dimensions. Bond et al. (2014) noted that some groups are less likely to be matched. These groups include young children, minorities, residents of group quarters, recent movers, low-income individuals, and the unemployed. Furthermore, Seeskin (2016) found that when households within these demographics do match they are more likely to link to missing information. For example, the odds of a household in poverty matching to available property tax data are 79 percent of that for households not in poverty. Likewise, the odds of rural households matching are 33.7 percent of that for non-rural households. The probability of matching to missing information decreases with higher educational attainment.

Research on the use energy and heating fuel data from household surveys and administrative records is scarce. When available, such studies tend to focus on fuel and energy costs rather than identifying specific energy types utilized. For example, Wiltse et al. (2014) compared administrative data from local utility companies in Alaska to the ACS to derive accurate statistics on energy usage, affordability and efficiency. These figures were then directly incorporated into state policy in the form of energy rebate programs and new construction standards. With respect to costs, they found differences in average energy costs between ACS and administrative records between very rural/remote areas and more populated geographies.

Lack of standardization of data across locations is one of many factors that contribute to discrepancies between survey data from energy supplier records. The U.S. Energy Information Administration (EIA) conducts the Residential Energy Consumption Survey (RECS) which collects data from householders, rental agents, and energy suppliers on energy-related structural characteristics, household energy usage, and energy consumption and cost for primary residences. O'Brien (2011) described the lack of knowledge by suppliers about *how* energy and fuel are used within a home. While the ACS inquires on the most used heating fuel within the housing unit, energy suppliers cannot with certainty say the heating fuel provided is the primary fuel used within the home. For example, a household may use a space heater rather than the central heating system fueled with natural gas; or natural gas usage is used primarily for cooking rather than heat. These occurrences likely vary by region. With the increased use of electronic bills and automatic payments, many the householders do see their monthly bill at all. This may result in inaccuracies in survey response. The EIA uses RECS data for important cost assessments and forecasting to manage national resource reserves.

3. ACS BACKGROUND

The housing section of the ACS includes questions on accessible utilities, included appliances, and heating fuel type associated with a housing unit. The remainder of this section describes the content of the ACS regarding these topics. The ACS is conducted via multiple data collection methods including: paper, internet, computer-assisted telephone interview (CATI), and computer-assisted personal interview (CAPI).[1] Descriptions using the paper data collection instrument for illustration follow.

3.1 Facilities

Question 8 inquires about the utilities and certain appliances available within the housing unit. It provides separate "yes" and "no" checkboxes to indicate the presence or absence of the following items. These items are: a) hot and cold running water, b) a flush toilet[2], c) a bathtub or shower, d) an indoor sink with a faucet, e) an indoor, non-portable stove or range, f) a refrigerator, and g) telephone service from which you can both make and receive calls (including cell phones).

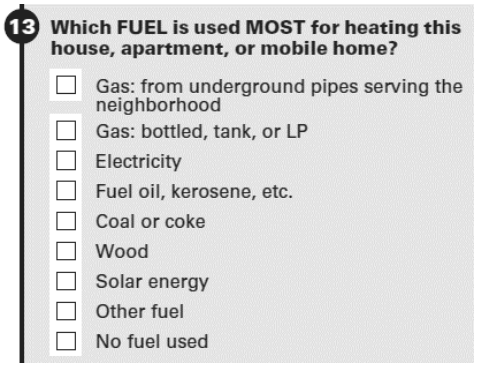

8 Does this house, apartment, or mobile home have –

	Yes	No
a. hot and cold running water?	☐	☐
b. a flush toilet?	☐	☐
c. a bathtub or shower?	☐	☐
d. a sink with a faucet?	☐	☐
e. a stove or range?	☐	☐
f. a refrigerator?	☐	☐
g. telephone service from which you can both make and receive calls? *Include cell phones.*	☐	☐

According to the ACS Information Guide, item *a* should be marked "yes," even if the unit has hot water only part of the time. Also, item *g* should be marked "yes" if there is a working phone present and someone receives service at the home, or someone has a cell phone that members of the household may use to make and receive calls. Discontinued phone service due to nonpayment should be marked "no."

3.2 Fuel Type

Question 13 asks, "Which FUEL is used MOST for heating this house, apartment, or mobile home?" Nine checkboxes are provided for the following responses: 1) Gas: from underground pipes serving the neighborhood, 2) Gas: bottled, tank, or LP, 3) Electricity, 4) Fuel oil, kerosene, etc., 5) Coal or coke, 6) Wood, 7) Solar energy, 8) Other fuel, and 9) No fuel used.


13 Which FUEL is used MOST for heating this house, apartment, or mobile home?

- ☐ Gas: from underground pipes serving the neighborhood
- ☐ Gas: bottled, tank, or LP
- ☐ Electricity
- ☐ Fuel oil, kerosene, etc.
- ☐ Coal or coke
- ☐ Wood
- ☐ Solar energy
- ☐ Other fuel
- ☐ No fuel used

The respondent should mark only one box indicating the heating fuel used most in the home. Residents of multi-unit apartment buildings should consult the owner, manager, or janitor to confirm the appropriate fuel type. The information guide defines solar energy as "a system that collects, stores, and distributes heat from the sun." Also, "other" fuel types may include energy sources such as purchased steam, fuel briquettes, and waste material.

---

[1] CATI was discontinued as a non-response follow-up mode in 2017; however it was in use in the 2014 ACS, the year of data which is used in the analysis in this report.

[2] The flush toilet question was discontinued in 2016, but, as it was on the 2014 ACS, it is used in this report.

4. RESEARCH QUESTIONS

Our research questions consist of the following:

1. How often does this administrative records source contain data that can be used to replace or supplement the respondent-provided response?

After matching 2014 ACS data to the 2014 vendor data by unique address identifiers called Master Address File IDs (MAFIDs), we address research question #1 by calculating the percentage of 2014 ACS households (both responding and vacant) that can be matched with vendor data records. Note that although the vendor data was delivered in 2014, not all records in the vendor data were listed in 2014.

2. To what extent is the construct in the administrative records in agreement with the construct as measured by the ACS questions?

The results section describes and compares key vendor data variables to the topical ACS items covered in this report in order to evaluate conceptual agreement and the suitability of the vendor data for ACS supplementation.

3. How often do the housing unit data from the administrative records source(s) agree with the responses from ACS householders by major subpopulation and housing characteristics?

For each topic, the results on response agreement are presented in two tables. The first set of results are detailed cross-tabulations of response values between the ACS and vendor data. The second table evaluates the relationship in responses across several demographic and housing characteristics. Next, the Methodology section discusses the criteria used for evaluating agreement in values within MAFID-linked data.

5. METHODOLOGY

MAFID-match Process and Data Description:

A vendor compiles MLS data that can be used to provide information on facilities and fuel type. They represent data delivered in 2014, the most recent year of data currently available to the Census Bureau at the time of the analysis, and are compared to edited values reported or edited in the 2014 ACS, including allocated responses.

In order to answer our research questions, we link each MLS record and ACS response by Master Address File Identification Number (MAFID) – a number associated with each record in the Master Address File (MAF). The MAF is a database containing the address or location description of every building (residential or non-residential) known to the Census Bureau, along with geographic information. ACS responses already have MAFIDs. We process the vendor data by appending MAFIDs to each address, where possible, and link on MAFID. We restrict our analysis to households where we link vendor data to an ACS unit by MAFID. Our analysis shows counts of ACS unweighted records.

Table 1: Vendor Data MAFID-Match Rate across ACS Households

	ACS data		Vendor data	
	All Units	Owner Occupied	All Units	Owner Occupied
Households/Housing Units/Records	2,270,415	1,291,614	50,119,977	11,621,909
Households with MAFID	2,270,415	1,291,614	46,571,852	3,992,235
Unique MAFIDs	-	-	20,662,779	2,525,138
Households with MAFID Matched ACS-Vendor Records	315,987	221,445	315,987	40,187
ACS-Vendor Match Rate	***13.92%***	***17.14%***	***1.53%***	***1.59%***

Source: 2014 1-year ACS Edited Data, MLS Data
Note: This table reports match rates for the 2,711 counties covered by the vendor data. Total housing units across all counties in the ACS in all counties = 2,322,722.

Table 1 provides the results of this matching process for ACS and vendor data. As mentioned previously, all households in the 2014 ACS have MAFIDs. Of the 50.1 million housing unit records in the 2014 MLS vendor data, 46.6 million (92.9 percent) have MAFIDs attached. Note, however, that due to the nature of the vendor data, an address can have multiple records, and so these matched records represent only 20.7 million unique addresses. The MLS data cover a wide range of geographies, with 2,711 counties (roughly 86 percent of counties) appearing in the data at least once. When we link ACS households in these 2,711 counties with the vendor data, we are only able to find 13.9 percent of all ACS housing units and 17.1 percent of owner-occupied ACS units in the MLS data. The relatively low match rate between the ACS and MLS data appears to be due primarily to poorer MLS coverage within the included areas rather than limited geographic coverage. For reference, appendix table 1A reports match rates for ACS respondents in all counties, not just the 2,711 counties reported in table 1.

This research compares the ACS respondent-provided data to the vendor data to analyze the potential for the vendor data to replace or supplement the facilities and fuel type questions on the ACS. The focus of this initial research is the linkage between the administrative records source and the ACS response data. We achieve this goal by summarizing the quality and distribution of the data on MAFID-matched housing units via match rates for the topics covered in this report.

Criteria for Evaluating Agreement:

The second goal of this research is to evaluate agreement in values between ACS and vendor data for MAFID-matched observations via a cross-tabulation comparison of the ACS and vendor data values. Clearly defined categorical response values are compared with vendor data based on ACS values outlined in the questionnaire.

We retain the same evaluation criteria when evaluating the coverage rates of the matched data over householder demographic and geographic categories. Additionally, we summarize the degree to which ACS and vendor data matches agree across these characteristics.

6. LIMITATIONS

There are some limitations that influence the quality of linked data and its suitability for ACS item replacement or imputation.

1. MLS records focus on housing units for sale as opposed to the entire universe of housing units.

2. MLS records are only available for particular housing markets at a certain point in time, which limits the availability of the data across the entire ACS sample.

These limitations relate to the coverage of MLS data. Given that MLS focuses mostly on houses for sale (only about 7 percent of records are for rental properties), it may miss a segment of the housing unit population that have not recently been listed for sale. Currently, the vendor is only able to provide MLS data for a specific number of markets, which also limits the applicability of this data. Although the counties in which at least some vendor data are available contained more than 98 percent of the 132.7 million housing units in the United States in 2014, only about 21 million unique addresses appear in the vendor data. It is possible that this issue could be ameliorated in the future as the vendor MLS data expands to cover more markets, even within the counties already represented in the data.

3. MLS records are collected for different objectives than ACS data, and the subject matter they capture may differ.

Since MLS data are derived from listings by real estate agents, the concepts captured may differ. Real estate listings are designed to sell houses, not to rigorously catalogue their characteristics. The features that are salient to potential homebuyers may differ from those that are of interest to data users. Moreover, many variables in the vendor data simply provide large blocks of text from these listings. If real estate agents use non-standardized language to describe features of the housing units they list, it can be difficult to identify characteristics of those units systematically. Typographical errors also complicate this process.

Additionally, note that throughout this report, results are reported for edited ACS data only, meaning differences between the administrative records and ACS data could arise from either responses or editing. However, as allocation is relatively rare for these ACS questions and the binding constraint in this case is the quality and coverage of the administrative records, the use of edited data should not fundamentally change the conclusions of this report.

7. RESULTS

1. *How often does this administrative records source contain data that can be used to replace or supplement the respondent provided response?*

We restrict our analysis to the 13.9 percent of ACS households that match to a vendor record by MAFID, resulting in 315,987 records total. However, note that even for linked records, the vendor data may still be missing information on the topics of interest. Table 2 displays the rate at which the vendor data provides non-missing and missing data for ACS records across the topics of this report.

Table 2: Match Rate by Non-missing and Missing Vendor Data among MAFID-matched Observations

ACS topic	Count, ACS records matched to vendor data	% of ACS households matched to non-missing vendor data	% of ACS households in covered counties matched to non-missing vendor data
Non-missing Vendor Data			
Facilities	315,657	13.59	13.90
Fuel Type	314,014	13.52	13.83
Missing Vendor Data			
Facilities	330	0.01	0.01
Fuel Type	1,973	0.08	0.09

Source: 2014 1-year ACS Edited Data, MLS Data

Note: Rows sum to total MAFID-matched observations. (Number of matched ACS households) = 315,987. (Total number of ACS households) = 2,322,722. (Total number of ACS households in counties covered by vendor data) = 2,270,415.

Both facilities and fuel type have low match rates to non-missing vendor data for ACS records. Consistent with Table 1, limiting the ACS data to counties that contain at least one MLS record does not lead to much improvement in the share of records that match to MLS data, because the vast majority of counties are represented in the MLS data. Again, the relatively low overall match rate appears to be driven by the limited number of units included in the MLS data within covered counties rather than by severely limited geographic coverage.

7.1 Facilities:

Part I: To what extent is the construct of administrative records in agreement with the construct as measured by the ACS questions?

Table 3 below provides a description of the agreement between MLS and ACS data on facilities. Each row provides answers to facilities questions in the ACS and each column provides a constructed measure of facilities derived from text strings in the MLS data.[3] Given that the overwhelming majority of ACS answers to these questions are "Yes," we focus on how well MLS data agrees on "No" responses. Looking through the different facilities types, MLS data do not agree with the vast majority of "No" responses by ACS respondents. The best agreement is for refrigerator. 4,365 of 315,987 (1.4 percent) matched ACS respondents responded "No," and of these, only 198 or 4.5 percent are coded as having no refrigerator in MLS. Therefore, it appears that there is very little agreement between ACS and MLS data on facilities when it comes to identifying the absence of particular items.

[3] See Table 6A in the appendix for a list of text strings used to construct the MLS facilities and fuel type measures used in the following tables.

Table 3: ACS Facilities (2014, edited) by Vendor Data, full sample

				MLS Data		
ACS		Unmatched	Matched	Missing	No	Yes
Hot and cold running water	Missing	0	0	0	0	0
	No	38,407	2,914	-	-	2,914
	Yes	1,968,328	313,073	328	461	312,284
Bathtub or shower	Missing	0	0	0	0	0
	No	25,024	1,499	0	13	1,486
	Yes	1,981,711	314,488	330	1,320	312,838
Refrigerator	Missing	0	0	0	0	0
	No	40,234	4,365	-	198	4,167
	Yes	1,966,501	311,622	328	7,157	304,137
Telephone service	Missing	189,656	24,481	25	514	23,942
	No	45,113	6,358	-	-	6,204
	Yes	1,771,966	285,148	301	6,382	278,465

Source: 2014 1-year ACS Edited Data, MLS Data
Note: "Missing", "No", and "Yes" values under MLS data sum to the value in the Matched column. "-" indicates a number was suppressed to avoid disclosure. Some numbers may be rounded to avoid disclosure.

Part II: How often do the housing unit data from the administrative records source(s) agree with the responses from ACS householders by major subpopulation and housing characteristics?

Before assessing agreement between the ACS and MLS data, it is worth considering coverage in more detail. The second column of Table 4 below shows differences in match rates across groups defined by householder/household characteristics. For example, the MLS coverage rates starts low for the youngest householders before increasing to its highest level for those 35-39 years of age and then declining to its lowest level for those who are at least 70 years of age. This pattern could reflect the competing influences of the composition of housing units contained in the MLS data and the market activities that generate the data. Housing units that appear in the MLS data were recently listed for sale. They are also predominantly single-family units. Younger householders may be more likely to live in a home that was recently listed, but that home is probably more likely to be a condominium or apartment than a single-family unit. Older householders may be less likely to live in a recently listed home, but their homes are probably more likely to be single-family units. Together, these patterns would suggest that middle-aged people are most likely to live in a recently listed single-family unit, which corresponds to the years for which the MLS data have the highest coverage rates. By way of context, note that the median age of single family owner householders who lived in a different home one year ago in the ACS is 45, which is consistent with the pattern in the MLS.

There are also substantial differences in coverage based on tenure and metropolitan status, with owned units and those in metropolitan areas matching to MLS records more often than rented or micropolitan and rural units. These differences are also consistent with the construction of the MLS data, as owner-occupied units are more likely to be listed than rental properties in the MLS, and the counties that lack any MLS data contain disproportionately few housing units, suggesting they are more likely to be rural.

Table 4 also summarizes agreement between MLS and ACS in columns 4-6. Here "agreement" is defined as the ACS and MLS data indicating the presence of exactly the same items—that the yes/no answers to all facilities questions are the same in both datasets. We present three statistics: agreement, disagreement and missing-ness, each of which is calculated as the percent of total matched records which agree/disagree/are missing. There is very little difference in agreement across demographic characteristics, but wide variation in agreement across geographic areas. While most of these agreement rates are high, as discussed above, the MLS and ACS data do not agree on the "No" responses. This limits the overall applicability of MLS data to the ACS process.

Table 4: ACS Facilities Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, full sample

	Matched Records	% of ACS Records Matched	Rate of Agreement		
			Agreement	Disagreement	Vendor is missing
Sex (of householder)					
Male	145,448	14.11	95.10	4.79	0.11
Female	146,058	13.55	95.49	4.41	0.10
Age (of householder)					
15-19	405	7.65	91.60	8.40	0.00
20-24	5,684	9.40	93.33	-	-
25-29	20,467	17.18	94.80	5.12	0.08
30-34	32,257	21.78	95.15	4.77	0.07
35-39	33,653	22.28	95.50	4.40	0.10
40-44	35,474	21.00	95.51	4.40	0.09
45-49	33,560	18.07	95.31	4.60	0.09
50-54	32,719	14.77	94.86	4.98	0.16
55-59	28,365	12.27	94.70	5.20	0.10
60-64	23,135	10.69	95.27	4.61	0.12
65-69	18,317	9.58	95.83	4.02	0.14
70 and over	27,470	6.71	96.51	3.37	0.12
Race (of householder)					
White alone	237,618	13.88	95.65	4.25	0.11
Black or African American alone	21,984	10.87	93.17	6.79	0.04
American Indian or Alaska Native alone	1,573	6.90	94.34	-	-
Asian alone	16,789	20.50	94.60	5.30	0.10
Native Hawaiian or Pacific Islander alone	250	10.89	95.20	4.80	0.00
Some Other Race alone	7,666	15.11	92.83	6.98	0.20
Two or More Races	5,626	15.23	94.31	5.56	0.12

Table 4: ACS Facilities Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, full sample (continued)

	Matched Records	% of ACS Records Matched	Rate of Agreement		
			Agreement	Disagreement	Vendor is missing
Ethnicity (of householder)					
Hispanic or Latino (of any race)	30,501	15.53	93.86	5.96	0.18
Not Hispanic or Latino	261,005	13.65	95.46	4.44	0.10
Place of birth (of householder)					
Native	245,818	13.30	95.43	4.46	0.10
Foreign born	45,587	17.51	94.54	5.35	0.11
Tenure					
Owned	237,024	15.98	95.70	4.20	0.10
Rented	54,482	8.71	93.53	6.34	0.13
MSA					
Metro	287,643	16.24	88.52	11.38	0.10
Micro	17,202	6.30	84.37	15.48	0.15
Other	11,142	3.99	77.73	22.12	0.14
State					
Alabama	1,111	2.96	82.09	17.91	0.00
Alaska	-	-	-	-	-
Arizona	3,810	8.53	86.90	13.10	0.00
Arkansas	1,759	7.82	91.02	8.70	0.28
California	41,340	19.04	90.03	9.73	0.24
Colorado	8,857	23.50	92.45	7.27	0.28
Connecticut	45	0.19	88.89	11.11	0.00
Delaware	2,322	31.52	86.95	13.05	0.00
District of Columbia	0	0.00			
Florida	31,640	25.97	80.98	18.89	0.13
Georgia	6,777	11.89	88.86	11.14	0.00
Hawaii	1,172	11.89	88.31	-	-
Idaho	1,814	15.71	91.95	-	-
Illinois	17,706	18.14	89.99	9.97	0.03
Indiana	161	0.33	86.96	13.04	0.00
Iowa	1,504	4.42	87.43	10.37	2.19
Kansas	-	-	-	-	-
Kentucky	1,078	3.16	90.17	9.83	0.00
Louisiana	2,207	7.07	87.72	-	-
Maine	3,000	17.01	87.10	12.90	0.00

Table 4: ACS Facilities Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, full sample (continued)

	Matched Records	% of ACS Records Matched	Rate of Agreement		
			Agreement	Disagreement	Vendor is missing
Maryland	91	0.23	89.01	10.99	0.00
Massachusetts	154	0.35	92.86	7.14	0.00
Michigan	14,963	14.82	83.61	-	-
Minnesota	19,910	27.42	89.98	9.45	0.07
Mississippi	881	4.70	86.15	13.85	0.00
Missouri	6,071	12.00	90.79	9.21	0.00
Montana	326	2.82	89.26	10.74	0.00
Nebraska	63	0.30	90.48	9.52	0.00
Nevada	7,262	39.71	87.14	-	-
New Hampshire	3,305	29.41	87.81	11.92	0.27
New Jersey	6,949	12.17	88.33	-	-
New Mexico	22	0.14	81.82	-	-
New York	14,868	10.72	88.96	-	-
North Carolina	12,493	18.14	86.94	12.96	0.10
North Dakota	25	0.26	64.00	36.00	0.00
Ohio	24,524	27.19	89.73	10.17	0.10
Oklahoma	8,733	18.82	86.50	13.42	0.08
Oregon	7,340	27.44	89.78	9.90	0.31
Pennsylvania	13,190	10.98	92.97	7.02	0.01
Rhode Island	2,031	29.78	91.24	8.76	0.00
South Carolina	1,017	3.08	81.81	-	-
South Dakota	67	0.67	91.04	8.96	0.00
Tennessee	905	2.05	91.60	8.40	0.00
Texas	24,320	16.56	87.77	12.20	0.03
Utah	-	-	-	-	-
Vermont	1,454	15.98	86.80	-	-
Virginia	5,358	9.80	74.90	25.01	0.09
Washington	1,701	3.56	88.71	-	-
West Virginia	159	1.03	89.31	10.69	0.00
Wisconsin	11,588	15.39	88.09	11.91	0.00
Wyoming	-	-	-	-	-

Source: 2014 1-year ACS Edited Data, MLS Data
Notes: "-" indicates a number was suppressed to avoid disclosure.

7.2 Heating Fuel Type:

Part I: To what extent is the construct of administrative records in agreement with the construct as measured by the ACS questions?

Table 5 displays a comparison of heating fuel type in the 2014 ACS and vendor data. Note that MLS can describe multiple fuel types, so one fuel type in ACS may be associated with one of multiple fuel types reported in the vendor data. Multiple fuel types can be reported for each record in the MLS, and, in fact, each individual ACS fuel type is associated with multiple MLS fuel types.

Table 5: ACS Heating Fuel Type (2014, edited) by Vendor Data, full sample

ACS Fuel Type	Matched ACS Records	MLS Fuel Type Data									
		Gas: from underground pipes	Gas: bottled, tank, LP	Electricity	Fuel oil, kerosene, etc.	Coal or coke	Wood	Solar energy	Other fuel	No fuel used	MLS data on fuel type not available
Gas: from underground pipes	172,955	125,117	171,685	171,062	171,685	171,684	50,826	1,196	171,813	54	1,134
Gas: bottled, tank, LP	13,295	4,721	13,188	13,167	13,188	13,188	4,250	100	13,195	-	-
Electricity	83,637	20,599	83,014	82,950	83,013	83,014	20,631	1,020	83,154	244	473
Fuel oil, kerosene, etc.	12,350	1,430	12,308	12,300	12,308	12,308	4,800	82	12,310	-	-
Coal or coke	131	27	130	130	130	130	52	-	130	-	-
Wood	49,10	1,094	4,876	4,866	4,876	4,876	2,183	-	4,878	-	32
Solar energy	252	66	251	251	251	251	70	50	251	0	-
Other fuel	1,190	327	1,174	1,174	1,174	1,174	367	-	1,174	0	-
No fuel used	2,786	599	2,711	2,711	2,711	2,711	431	62	2,733	13	53
ACS data on fuel type not available	24,481	9,650	24,324	24,277	24,324	24,323	5,899	181	24,353	144	124

Source: 2014 1-year ACS Edited Data, MLS Data

Notes: "-" indicates a number was suppressed to avoid disclosure. Some numbers may be rounded to avoid disclosure. Multiple fuel types can be reported in the MLS, so the columns in this table are not mutually exclusive.

Part II: How often do the housing unit data from the administrative records source(s) agree with the responses from ACS householders by major subpopulation and housing characteristics?

Table 6 provides coverage and agreement rates for fuel type across demographic characteristics and geographies. The second column reports the same coverage rates as Table 4, which are based on record matches across datasets. MLS records that match successfully to the ACS are missing fuel type or facilities information in only a limited number of cases, as shown in Table 2, so differences in available information across ACS topics are minimal. For fuel type, virtually no observations would meet the agreement standard applied to facilities (because almost all units list multiple fuel types in the MLS data). Instead of producing row after row of zeroes, Table 6 reports how often the ACS reported or edited primary fuel type is included in the list of available fuel types derived from the MLS data. As with facilities, there are substantial differences in agreement across geographies or demographic characteristics.

Table 6: ACS Heating Fuel Type Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, full sample

	Matched Records	% of ACS Records Matched	Rate of Agreement[4]		
			Agreement	Disagreement	Vendor is missing
Sex (of householder)					
Male	145,448	14.11	80.39	18.93	0.68
Female	146,058	13.55	82.28	17.12	0.59
Age (of householder)					
15-19	405	7.65	80.49	-	-
20-24	5,684	9.40	82.83	16.61	0.56
25-29	20,467	17.18	82.92	16.65	0.43
30-34	32,257	21.78	82.40	17.16	0.43
35-39	33,653	22.28	81.31	18.14	0.55
40-44	35,474	21.00	80.68	18.79	0.54
45-49	33,560	18.07	80.30	19.12	0.58
50-54	32,719	14.77	80.46	18.82	0.72
55-59	28,365	12.27	80.59	18.62	0.80
60-64	23,135	10.69	81.13	18.07	0.80
65-69	18,317	9.58	82.07	17.09	0.84
70 and over	27,470	6.71	82.30	16.92	0.78
Race (of householder)					
White alone	237,618	13.88	82.63	16.76	0.62
Black or African American alone	21,984	10.87	85.62	13.94	0.44
American Indian or Alaska Native alone	1,573	6.90	77.37	22.06	0.57
Asian alone	16,789	20.50	67.10	31.74	1.16
Native Hawaiian or Pacific Islander alone	250	10.89	65.20	-	-
Some Other Race alone	7,666	15.11	64.92	34.53	0.55
Two or More Races	5,626	15.23	76.84	22.41	0.75

[4]Denominator in the rates of agreement calculations is the value in the Matched Records column.

Table 6: ACS Heating Fuel Type coverage and match rates by householder and geography characteristics (edited 2014 ACS) for data linked by MAFID to Vendor Data, full sample (continued)

	Matched Records	% of ACS Records Matched	Rate of Agreement		
			Agreement	Disagreement	Vendor is missing
Ethnicity (of householder)					
Hispanic or Latino (of any race)	30,501	15.53	70.15	29.38	0.47
Not Hispanic or Latino	261,005	13.65	82.65	16.70	0.65
Place of birth (of householder)					
Native	245,818	13.30	82.85	16.52	0.63
Foreign born	45,587	17.51	73.17	26.14	0.68
Tenure					
Owned	237,024	15.98	81.68	17.69	0.63
Rented	54,482	8.71	79.86	19.51	0.63
MSA					
Metro	287,643	16.24	75.59	23.79	0.62
Micro	17,202	6.30	70.20	29.10	0.70
Other	11,142	3.99	68.19	31.11	0.70
State					
Alabama	1,111	2.96	69.76	29.34	0.90
Alaska	-	-	-	-	-
Arizona	3,810	8.53	75.88	22.28	1.84
Arkansas	1,759	7.82	72.14	18.65	9.21
California	41,340	19.04	37.51	61.57	0.92
Colorado	8,857	23.50	69.29	30.36	0.35
Connecticut	45	0.19	82.22	17.78	0.00
Delaware	2,322	31.52	83.72	16.28	0.00
District of Columbia	0	0.00			
Florida	31,640	25.97	79.21	20.46	0.33
Georgia	6,777	11.89	82.00	-	-
Hawaii	1,172	11.89	36.43	58.45	5.12
Idaho	1,814	15.71	84.95	14.44	0.61
Illinois	17,706	18.14	77.93	21.96	0.11
Indiana	161	0.33	87.58	12.42	0.00
Iowa	1,504	4.42	86.70	7.85	5.45
Kansas	-	-	-	-	-
Kentucky	1,078	3.16	90.82	-	-
Louisiana	2,207	7.07	74.81	24.88	0.32
Maine	3,000	17.01	82.43	-	-

Table 6: ACS Heating Fuel Type coverage and match rates by householder and geography characteristics (edited 2014 ACS) for data linked by MAFID to Vendor Data, full sample (continued)

	Matched Records	% of ACS Records Matched	Rate of Agreement		
			Agreement	Disagreement	Vendor is missing
Maryland	91	0.23	86.81	13.19	0.00
Massachusetts	154	0.35	88.96	11.04	0.00
Michigan	14,963	14.82	86.95	12.08	0.97
Minnesota	19,910	27.42	83.65	14.82	1.03
Mississippi	881	4.70	73.44	25.65	0.91
Missouri	6,071	12.00	89.23	10.67	0.10
Montana	326	2.82	90.18	-	-
Nebraska	63	0.30	87.30	12.70	0.00
Nevada	7,262	39.71	71.76	27.47	0.77
New Hampshire	3,305	29.41	79.36	20.15	0.48
New Jersey	6,949	12.17	84.10	15.74	0.16
New Mexico	22	0.14	72.73	27.27	0.00
New York	14,868	10.72	84.32	15.18	0.50
North Carolina	12,493	18.14	83.57	16.10	0.33
North Dakota	25	0.26	60.00	-	-
Ohio	24,524	27.19	87.06	12.24	0.70
Oklahoma	8,733	18.82	77.60	22.01	0.39
Oregon	7,340	27.44	84.37	15.01	0.61
Pennsylvania	13,190	10.98	89.40	-	-
Rhode Island	2,031	29.78	85.38	-	-
South Carolina	1,017	3.08	75.71	-	-
South Dakota	67	0.67	91.04	8.96	0.00
Tennessee	905	2.05	86.08	13.92	0.00
Texas	24,320	16.56	86.22	13.46	0.32
Utah	-	-	-	-	-
Vermont	1,454	15.98	78.27	-	-
Virginia	5,358	9.80	83.78	15.45	0.77
Washington	1,701	3.56	85.89	12.40	1.70
West Virginia	159	1.03	17.61	82.39	0.00
Wisconsin	11,588	15.39	43.89	55.66	0.45
Wyoming	-	-	-	-	-

Source: 2014 1-year ACS Edited Data, MLS Data

7.3 Vacant Housing Units:

In this section, we document the coverage and availability of facilities and fuel type data for vacant units in the ACS. Table 7 illustrates that eleven percent of vacant ACS units can be linked

to MLS data on facilities and fuel type, which is only slightly lower than the match rate for non-vacant units shown previously.

Table 7: Match Rate by Non-missing and Missing Vendor Data among MAFID-matched Observations, vacant unit sample

ACS topic	Count, ACS records matched to non-missing vendor data	% of ACS households matched to non-missing vendor data	Count, ACS records matched to missing vendor data	% of ACS households matched to missing vendor data
Facilities	24,456	11.42	25	0.01
Fuel Type	24,357	11.37	124	0.06

Source: 2014 1-year ACS Edited Data, MLS Data
Note: Rows sum to total MAFID-matched obs. (Number of matched ACS units) = 24,481. (Total number of vacant ACS units) = 214,137.

As mentioned above, the MLS data cover the vast majority of counties with what appears to be relatively little depth. As with occupied units, low match rates among vacant units are likely driven by the limited number of records available in covered areas rather than limited geographic coverage.

Part I: To what extent is the construct of administrative records in agreement with the construct as measured by the ACS questions?

Table 8 displays the relationship between facilities in MLS and ACS data. Similar to what was found for non-vacant units, the MLS data provides a "Yes" for most facilities and does a poor job of predicting "No" responses. For example, out of the 3,886 vacant units in ACS coded as not having a refrigerator, only 174 have no refrigerator in MLS. This implies that there is little information in the MLS data that predicts ACS responses for vacant units. Note that, since telephone information is not collected for vacant units, all ACS observations are treated as missing for this variable.

Table 8: ACS Facilities (2014, edited) by Vendor Data, vacant unit sample

				MLS Data		
ACS		Unmatched	Matched	Missing	No	Yes
Hot and cold running water	Missing	0	0	0	0	0
	No	32,079	2,445	-	-	2,445
	Yes	157,577	22,036	23	37	21,999
Bathtub or shower	Missing	0	0	0	0	0
	No	19,912	1,116	0	11	1,105
	Yes	169,744	23,365	25	67	23,273
Refrigerator	Missing	0	0	0	0	0
	No	34,193	3,886	-	174	3,712
	Yes	155,463	20,595	23	363	20,209
Telephone service	Missing	189,656	24,481	25	514	23,942
	No	0	0	0	0	0
	Yes	0	0	0	0	0

Source: 2014 1-year ACS Edited Data, MLS Data
Note: "Missing", "No", and "Yes" values under MLS data sum to the value in the Matched column. "-" indicates a number was suppressed to avoid disclosure. Numbers may be rounded to avoid disclosure.

ACS collects no data on fuel type for vacant units, so there is no way to compare ACS responses to MLS data for these units. Therefore, Table 9 presents what fuel type data is available for vacant housing units in the ACS. Here, we see that the majority of vacant units have a number of fuel types coded from the MLS data.

Table 9: ACS Heating Fuel Type (2014, edited) by Vendor Data, vacant unit sample

ACS Fuel Type	Matched ACS Records	MLS Fuel Type Data									
		Gas: from under-ground pipes	Gas: bottled, tank, LP	Electricity	Fuel oil, kerosene, etc.	Coal or coke	Wood	Solar energy	Other fuel	No fuel used	MLS data on fuel type not available
ACS data on fuel type not available	24,481	9,650	24,324	24,277	24,324	24,323	5,899	181	24,353	144	124

Source: 2014 1-year ACS Edited Data, MLS Data

Part II: How often do the housing unit data from the administrative records source(s) agree with the responses from ACS householders by major subpopulation and housing characteristics?

Table 10 displays the coverage and agreement for facilities in vacant ACS units. Since vacant units have no information for demographic characteristics, these rows are omitted. Looking across geography, the coverage rate is again higher in metropolitan areas, and there are differences in the availability and coverage of MLS data across geographies. Although there are no predictable geographical patterns in rates of agreement between the ACS and MLS data, there is substantial variability, agreement rates varying from less than 50 percent (Michigan) to 100 percent (Hawaii).

Table 10: ACS Facilities Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, vacant unit sample

	Matched Records	% of ACS Records Matched	Rate of Agreement[5]		
			Agreement	Disagreement	vendor is missing
MSA					
Metro	20,384	16.36	78.75	21.15	0.10
Micro	2,081	5.90	84.62	-	-
Other	2,016	3.71	84.87	-	-
State					
Alabama	178	4.31	81.46	18.54	0.00
Alaska	0	0.00			
Arizona	389	7.01	89.97	10.03	0.00
Arkansas	111	4.55	62.16	37.84	0.00
California	2,104	17.06	76.66	22.91	0.43
Colorado	467	15.82	93.15	-	-
Connecticut	-	-	-	-	-
Delaware	249	25.00	92.77	7.23	0.00
District of Columbia	0	0.00			
Florida	4,908	32.16	93.19	-	-
Georgia	591	11.04	77.33	22.67	0.00
Hawaii	104	9.39	100.00	0.00	0.00
Idaho	97	6.90	85.57	-	-
Illinois	1,161	18.75	72.27	27.73	0.00
Indiana	-	-	-	-	-
Iowa	55	3.03	67.27	32.73	0.00
Kansas	0	0.00			
Kentucky	54	2.04	66.67	33.33	0.00
Louisiana	151	4.47	66.89	33.11	0.00
Maine	305	7.01	92.13	7.87	0.00
Maryland	-	-	-	-	-
Massachusetts	-	-	-	-	-
Michigan	1,273	10.03	44.70	55.30	0.00
Minnesota	1,093	15.43	86.46	13.54	0.00
Mississippi	90	4.42	85.56	14.44	0.00
Missouri	386	8.05	72.80	27.20	0.00
Montana	20	1.05	80.00	-	-
Nebraska	-	-	-	-	-
Nevada	675	35.06	91.11	8.89	0.00
New Hampshire	324	21.26	94.14	5.86	0.00

[5] Denominator in the rates of agreement calculations is the value in the Matched Records column.

Table 10: ACS Facilities Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, vacant unit sample (continued)

	Matched Records	% of ACS Records Matched	Rate of Agreement		
			Agreement	Disagreement	Vendor is missing
New Jersey	658	14.38	83.74	16.26	0.00
New Mexico	-	-	-	-	-
New York	1,081	8.67	84.37	-	-
North Carolina	1,279	16.91	87.49	-	-
North Dakota	-	-	-	-	-
Ohio	1,495	24.55	56.86	-	-
Oklahoma	823	11.78	64.03	35.97	0.00
Oregon	362	20.14	86.19	-	-
Pennsylvania	606	5.59	67.16	32.84	0.00
Rhode Island	137	25.90	81.75	18.25	0.00
South Carolina	162	4.37	92.59	7.41	0.00
South Dakota	-	-	-	-	-
Tennessee	59	1.70	88.14	11.86	0.00
Texas	1,282	10.22	71.68	-	-
Utah	0	0.00			
Vermont	167	9.85	91.62	8.38	0.00
Virginia	392	11.40	70.41	-	-
Washington	71	2.25	95.77	-	-
West Virginia	11	0.54	72.73	-	-
Wisconsin	1,079	13.39	84.99	15.01	0.00
Wyoming	-	-	-	-	-

Source: 2014 1-year ACS Edited Data, MLS Data
Notes: "-" indicates a number was suppressed to avoid disclosure.

8. CONCLUSIONS

We linked MLS data on fuel and facilities type to ACS microdata to investigate the quality and coverage of this vendor data and the degree to which this vendor data could potentially be used to supplement or replace ACS questions. We find very low match rates, overall, but high rates of agreement within the matched records. After linking the 2014 ACS to the vendor data by MAFID, we find 13.9 percent of ACS households have a match within the MLS data, in counties with MLS coverage. Across ACS topics, the rates at which ACS records match to missing information within vendor data are very low at less than one percent. When ACS records match to non-missing vendor data, the rate of agreement of linked values for facilities questions is 95.3 percent, but the MLS data correctly identify very few negative (absence) responses. The vendor data cannot identify a single primary fuel type, but it includes the fuel type listed by the ACS respondent 81 percent of the time.

Across demographic and geographic characteristics, ACS households that are owner-occupied, located in metropolitan areas, and have householders who are neither very young nor very old are more likely to match to MLS records. There is little variation in the agreement rates between

ACS and MLS vendor data across the various characteristics outlined in the tables above. Within a subset of vacant units, coverage and agreement rates display similar results. One exception is the evaluation of agreement in fuel type. MLS data details available fuels within the unit in the absence of ACS information, which results in multiple fuel types listed for virtually all linked MLS records. Therefore, the agreement rate for this particular topic is zero. However, there are several non-zero rates describing coverage and agreement across geography. In particular, coverage is significantly higher in non-rural areas.

There are some limitations that influence the quality of linked data and its suitability for ACS item replacement or imputation. Due to differences in the objectives and methods of collecting the data, it is difficult for vendor data to conceptually align with the ACS questionnaire. In addition, MLS records often focus on housing units for sale as opposed to the entire universe of housing units, and have very little coverage of rental units. Additionally, MLS records are only available for particular housing markets at this point in time, which limits the availability of the data across the entire ACS sample. These differences raise concerns that characteristics that vary by housing unit may not align with the same information collected by MLS records. This is of particular risk among renters in multi-family buildings. Some of these limitations may be ameliorated in future deliveries of MLS vendor data, if they cover a larger number of housing units.

9. REFERENCES

Bond, B., Brown, J., Luque, A., O'Hara, A. (2014). "The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey", CARRA Working Paper Series, Working Paper #2014-08.

Brummet, Q. O. (2014). Comparison of survey, federal, and commercial address quality. CARRA Working Paper #2014-06. Washington, D.C.: U.S. Census Bureau.

Chappell, G., Obenski, S. (2014). "Fiscal Year 2014 Content Review Results", American Community Survey Fiscal Year 2014 Content Review.

Luque, A., Bhaskar, R. (2014). "2010 American Community Survey Match Study", CARRA Working Paper Series, Working Paper #2014-03.

O'Brien, E. (2011). Energy Records in lieu of Bills and Self Reports: Optimizing data utility, quality and burden in the Residential Energy Consumption Survey (RECS). Paper presented at the Committee on National Statistics—Household Survey Producers Workshop, Washington, DC.

Ruggles, P. (2015). "Review of Administrative Data Sources Relevant to the American Community Survey", Prepared for the U.S. Census Bureau, January 31.

Seeskin, Z. H. (2016). "Evaluating the Use of Commercial Data to Improve Survey Estimates of Property Taxes", CARRA Working Paper #2016-06. Washington, D.C.: U.S. Census Bureau.

U.S. Census Bureau. (2015). "Agility in Action: A Snapshot of Enhancements to the American Community Survey", https://www.census.gov/content/dam/Census/programs-surveys/acs/operations-and-administration/2015-16-survey-enhancements/Agility_in_Action_V1.1.pdf.

Wiltse, N., Madden, D., Valentine, B., Stevens, V. (2014). "2013 Alaska Housing Assessment", Cold Climate Housing Research Center. Prepared for: Alaska Housing Finance Corporation.

APPENDIX

Table 11: Unrestricted ACS Match Rates

	ACS		
	Full sample	Owner Occupied	Vacant
Households/Housing Units/Records	2,322,722	1,320,918	214,137
Households with MAFID Match	315,987	221,445	24,481
Match Rate Across ACS Households	***13.60%***	***16.76%***	***11.43%***

Source: 2014 1-year ACS Edited Data, MLS Data

Table 12: Match Rate by Non-missing and Missing Vendor Data among MAFID-matched Observations, single family, owner-occupied sample

ACS topic	Count, ACS records matched to non-missing vendor data	% of ACS households matched to non-missing vendor data	Count, ACS records matched to missing vendor data	% of ACS households matched to missing vendor data
Facilities	221,236	16.75	209	0.02
Fuel Type	220,077	16.66	1,368	0.10

Source: 2014 1-year ACS Edited Data, MLS Data

Note: Rows sum to total MAFID-matched obs. (Number of ACS households) = 221,445

Table 13: ACS Facilities (2014, edited) by Vendor Data, single family, owner-occupied sample

				MLS Data		
ACS		Unmatched	Matched	Missing	No	Yes
Hot and cold running water	Missing	0	0	0	0	0
	No	3,175	282	0	0	282
	Yes	1,096,298	221,163	209	292	220,662
Bathtub or shower	Missing	0	0	0	0	0
	No	2,488	238	0	0	2,726
	Yes	1,096,985	221,207	209	830	220,168
Refrigerator	Missing	0	0	0	0	0
	No	2,184	272	0	11	261
	Yes	1,097,289	221,173	209	4,941	216,023
Telephone service	Missing	0	0	0	0	0
	No	15,957	4,133	-	94	4,039
	Yes	1,083,516	217,312	207	4,711	212,394

Source: 2014 1-year ACS Edited Data, MLS Data

Note: "Missing", "No", and "Yes" values under MLS data sum to the value in the Matched column. "-" indicates a statistic was suppressed to avoid disclosure. Numbers in table may also be rounded to avoid disclosure.

Table 14: ACS Facilities Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, single family, owner-occupied sample

	Matched Records	% of ACS Records Matched	Rate of Agreement[6]		
			Agreement	Disagreement	Vendor is missing
Sex (of householder)					
Male	114,207	16.71	95.48	4.42	0.10
Female	107,238	16.82	96.02	3.89	0.09
Age (of householder)					
15-19	136	22.63	89.71	10.29	0.00
20-24	2,238	28.83	93.70	6.30	0.00
25-29	12,722	35.11	95.21	4.72	0.07
30-34	22,971	34.39	95.75	4.19	0.06
35-39	25,053	30.81	96.09	3.84	0.07
40-44	27,386	26.70	95.98	3.93	0.09
45-49	26,680	21.99	95.66	4.27	0.07
50-54	26,379	17.44	95.23	4.64	0.14
55-59	22,928	14.01	95.11	4.80	0.09
60-64	18,646	11.89	95.74	4.17	0.10
65-69	14,801	10.54	96.08	3.78	0.14
70 and over	21,505	7.36	96.75	3.12	0.13
Race (of householder)					
White alone	186,105	16.29	95.98	3.92	0.10
Black or African American alone	12,908	15.04	94.08	-	-
American Indian or Alaska Native alone	1,054	9.09	95.07	-	-
Asian alone	12,696	28.42	95.10	4.84	0.06
Native Hawaiian or Pacific Islander alone	134	12.85	97.76	-	-
Some Other Race alone	4,842	26.52	92.94	6.86	0.21
Two or More Races	3,706	21.47	95.22	-	-

[6] Denominator in the rates of agreement calculations is the value in the Matched Records column.

Table 14: ACS Facilities Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, single family, owner-occupied sample (continued)

	Matched Records	% of ACS Records Matched	Rate of Agreement		
			Agreement	Disagreement	Vendor is missing
Ethnicity (of householder)					
Hispanic or Latino (of any race)	19,907	23.20	94.27	5.56	0.17
Not Hispanic or Latino	201,538	16.32	95.88	4.03	0.09
Place of birth (of householder)					
Native	188,880	15.83	95.87	4.03	0.10
Foreign born	32,483	25.50	94.98	4.93	0.09
Tenure					
Owned	221,445	16.76	95.74	4.17	0.09
MSA					
Metro	202,054	20.03	95.71	4.20	0.09
Micro	12,075	7.62	96.34	3.49	0.17
Other	7,316	4.76	95.50	4.37	0.12
State					
Alabama	766	3.61	98.04	1.96	0.00
Alaska	-	-	-	-	-
Arizona	2,499	10.92	97.44	2.56	0.00
Arkansas	1,324	10.77	97.51	2.11	0.38
California	27,350	24.84	95.54	4.28	0.18
Colorado	6,566	29.54	97.76	1.93	0.30
Connecticut	33	0.23	93.94	-	-
Delaware	1,736	39.63	97.87	2.13	0.00
District of Columbia	0.00	0.00			
Florida	16,523	28.04	96.45	3.38	0.17
Georgia	4,530	14.51	97.40	2.60	0.00
Hawaii	441	9.64	97.05	-	-
Idaho	1,335	20.07	96.93	3.07	0.00
Illinois	11,559	19.94	96.39	-	-
Indiana	129	0.41	93.02	6.98	0.00
Iowa	1,241	5.20	91.38	6.37	2.26
Kansas	-	-	-	-	-
Kentucky	809	4.06	94.68	5.32	0.00
Louisiana	1,687	10.20	95.38	-	-
Maine	2,191	23.93	97.26	2.74	0.00
Maryland	75	0.30	97.33	-	-
Massachusetts	122	0.53	96.72	-	-
Michigan	10,934	17.27	92.35	-	-

Table 14: ACS Facilities Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, single family, owner-occupied sample (continued)

	Matched Records	% of ACS Records Matched	Rate of Agreement		
			Agreement	Disagreement	Vendor is missing
Minnesota	16,178	32.45	96.02	3.90	0.08
Mississippi	618	5.99	96.76	3.24	0.00
Missouri	4,571	15.03	97.40	2.60	0.00
Montana	267	4.35	94.38	5.62	0.00
Nebraska	51	0.36	94.12	-	-
Nevada	4,069	46.72	96.41	3.59	0.00
New Hampshire	2,443	36.63	97.50	2.17	0.33
New Jersey	5,202	15.99	97.81	-	-
New Mexico	12	0.15	91.67	-	-
New York	11,119	17.52	96.84	-	-
North Carolina	8,489	22.67	97.20	2.70	0.11
North Dakota	23	0.39	65.22	34.78	0.00
Ohio	18,577	33.15	96.23	3.71	0.06
Oklahoma	6,097	25.33	96.01	3.89	0.10
Oregon	5,312	36.65	95.20	4.54	0.26
Pennsylvania	10,459	13.65	97.94	2.06	0.00
Rhode Island	1,517	42.47	98.22	1.78	0.00
South Carolina	677	3.81	98.08	-	-
South Dakota	53	0.86	96.23	-	-
Tennessee	672	2.56	98.81	1.19	0.00
Texas	18,163	21.83	92.34	-	-
Utah	-	-	-	-	--
Vermont	1,078	21.03	98.79	1.21	0.00
Virginia	3,815	11.33	81.60	-	-
Washington	1,263	4.68	93.27	-	-
West Virginia	124	1.42	95.97	4.03	0.00
Wisconsin	8,733	18.10	97.48	2.52	0.00
Wyoming	-	-	-	-	-

Source: 2014 1-year ACS Edited Data, MLS Data

Notes: "-" indicates a statistic was suppressed to avoid disclosure.

Table 15: ACS Heating Fuel Type (2014, edited) by Vendor Data, single family, owner-occupied sample

ACS Fuel Type	Matched ACS Records	MLS Fuel Type Data									
		Gas: from underground pipes	Gas: bottled, tank, LP	Electricity	Fuel oil, kerosene, etc.	Coal or coke	Wood	Solar energy	Other fuel	No fuel used	MLS data on fuel type not available
Gas: from underground pipes	138,126	101,782	137,126	136,625	137,126	137,126	42,796	1,085	137,218	137,218	901
Gas: bottled, tank, LP	11,011	3,839	10,922	10,907	10,922	10,922	3,711	90	10,927	10,927	83
Electricity	55,116	14,176	54,733	54,692	54,733	54,733	15,261	884	54,711	54,811	301
Fuel oil, kerosene, etc.	10,368	1,096	10,339	10,333	10,339	10,339	4,221	80	10,340	10,340	28
Coal or coke	117	23	116	116	116	116	50	-	116	116	-
Wood	4,226	958	4,198	4,188	4,198	4,198	1,934	64	4,199	4,199	27
Solar energy	226	56	225	225	225	225	62	49	225	225	-
Other fuel	887	220	876	875	876	876	291	10	876	876	11
No fuel used	1,368	277	1,337	1,334	1,337	1,337	266	52	1,353	1,353	15
ACS data on fuel type not available	-	-	-	-	-	-	-	-	-	-	-

Source: 2014 1-year ACS Edited Data, MLS Data
Notes: "-" indicates a statistic was suppressed to avoid disclosure. Multiple fuel types can be reported in the MLS, so the columns in this table are not mutually exclusive.

Table 16: ACS Heating Fuel Type Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, single family, owner-occupied sample

	Matched Records	% of ACS Records Matched	Rate of Agreement[7]		
			Agreement	Disagreement	Vendor is missing
Sex (of householder)					
Male	114,207	16.71	80.61	18.73	0.66
Female	107,238	16.82	82.66	16.76	0.58
Age (of householder)					
15-19	136	22.63	82.35	17.65	0.00
20-24	2,238	28.83	82.35	17.11	0.54
25-29	12,722	35.11	83.38	16.26	0.35
30-34	22,971	34.39	83.05	16.59	0.36
35-39	25,053	30.81	82.03	17.51	0.45
40-44	27,386	26.70	81.00	18.48	0.51
45-49	26,680	21.99	80.45	18.96	0.59
50-54	26,379	17.44	80.71	18.56	0.72
55-59	22,928	14.01	80.79	18.41	0.80
60-64	18,646	11.89	81.30	17.94	0.76
65-69	14,801	10.54	82.30	16.86	0.84
70 and over	21,505	7.36	82.39	16.79	0.81
Race (of householder)					
White alone	186,105	16.29	82.90	16.50	0.61
Black or African American alone	12,908	15.04	85.85	13.69	0.46
American Indian or Alaska Native alone	1,054	9.09	78.94	20.49	0.57
Asian alone	12,696	28.42	66.96	32.05	0.99
Native Hawaiian or Pacific Islander alone	134	12.85	58.96	-	-
Some Other Race alone	4,842	26.52	63.94	35.58	0.48
Two or More Races	3,706	21.47	76.71	22.58	0.70

[7] Denominator in the rates of agreement calculations is the value in the Matched Records column.

Table 16: ACS Heating Fuel Type Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, single family, owner-occupied sample (continued)

	Matched Records	% of ACS Records Matched	Rate of Agreement		
			Agreement	Disagreement	Vendor is missing
Ethnicity (of householder)					
Hispanic or Latino (of any race)	19,907	23.20	69.35	30.20	0.46
Not Hispanic or Latino	201,538	16.32	82.82	16.55	0.63
Place of birth (of householder)					
Native	188,880	15.83	83.08	16.31	0.61
Foreign born	32,483	25.50	73.01	26.35	0.65
Tenure					
Owned	221,445	16.76	81.61	17.78	0.62
MSA					
Metro	202,054	20.03	81.67	17.72	0.61
Micro	12,075	7.62	79.70	19.60	0.70
Other	7,316	4.76	83.08	16.31	0.62
State					
Alabama	766	3.61	83.68	-	-
Alaska	-	-	-	-	-
Arizona	2,499	10.92	86.67	11.48	1.84
Arkansas	1,324	10.77	76.66	12.92	10.42
California	27,350	24.84	36.58	62.66	0.76
Colorado	6,566	29.54	72.78	26.85	0.37
Connecticut	33	0.23	87.88	-	-
Delaware	1,736	39.63	93.84	6.16	0.00
District of Columbia	0	0.00			
Florida	16,523	28.04	93.97	5.63	0.39
Georgia	4,530	14.51	90.26	-	-
Hawaii	441	9.64	37.19	61.68	1.13
Idaho	1,335	20.07	89.74	9.66	0.60
Illinois	11,559	19.94	82.57	17.32	0.11
Indiana	129	0.41	90.70	9.30	0.00
Iowa	1,241	5.20	90.41	4.59	5.00
Kansas	-	-	-	-	-
Kentucky	809	4.06	96.04	3.96	0.00
Louisiana	1,687	10.20	79.85	19.80	0.36
Maine	2,191	23.93	91.19	-	-
Maryland	75	0.30	94.67	-	-
Massachusetts	122	0.53	91.80	8.20	0.00
Michigan	10,934	17.27	94.90	3.92	1.18

Table 16: ACS Heating Fuel Type Coverage and Match Rates by Householder and Geography Characteristics (edited 2014 ACS) for Data Linked by MAFID to Vendor Data, single family, owner-occupied sample (continued)

	Matched Records	% of ACS Records Matched	Rate of Agreement		
			Agreement	Disagreement	Vendor is missing
Minnesota	16,178	32.45	88.86	10.01	1.13
Mississippi	618	5.99	81.39	18.45	0.16
Missouri	4,571	15.03	95.80	-	-
Montana	267	4.35	97.38	-	-
Nebraska	51	0.36	94.12	-	-
Nevada	4,069	46.72	75.79	23.27	0.93
New Hampshire	2,443	36.63	87.88	11.58	0.53
New Jersey	5,202	15.99	93.23	6.63	0.13
New Mexico	12	0.15	83.33	-	-
New York	11,119	17.52	93.09	6.52	0.39
North Carolina	8,489	22.67	93.59	6.09	0.32
North Dakota	23	0.39	60.87	-	-
Ohio	18,577	33.15	93.21	6.09	0.69
Oklahoma	6,097	25.33	86.53	13.04	0.43
Oregon	5,312	36.65	88.29	11.09	0.62
Pennsylvania	10,459	13.65	93.86	-	-
Rhode Island	1,517	42.47	92.16	-	-
South Carolina	677	3.81	89.07	-	-
South Dakota	53	0.86	92.45	-	-
Tennessee	672	2.56	91.67	8.33	0.00
Texas	18,163	21.83	91.60	8.10	0.29
Utah	-	-	-	-	-
Vermont	1,078	21.03	87.66	-	-
Virginia	3,815	11.33	90.98	8.39	0.63
Washington	1,263	4.68	89.07	9.18	1.74
West Virginia	124	1.42	19.35	80.65	0.00
Wisconsin	8,733	18.10	46.52	53.02	0.46
Wyoming	-	-	-	-	-

Source: 2014 1-year ACS Edited Data, MLS Data
Notes: "-" indicates a statistic was suppressed to avoid disclosure

Table 17: Text Strings used to Create Facilities and Fuel Type Measures in MLS Data

Facilities	Fuel Type
Hot and cold running water	*Gas: from underground pipes*
sink	gas
tub	*Gas: bottled, tank, or LP*
shower	liquefied petroleum
pool	lpg
water	propane
Bathtub or shower	butane
tub	*Fuel oil, kerosene, etc.*
shower	oil
Refrigerator	kerosene
refrigerator	*Coal or coke*
fridge	coal
Telephone service	coke
telephone	*Wood*
phone	wood
	Solar energy
	solar
	Other fuel
	steam
	briquettes
	geothermal
	compost
	wind
	biomass
	hydropower

Source: 2014 1-year ACS Edited Data, MLS Data
Note: All relevant text-based MLS variables are converted to lower case before they are searched for these strings.