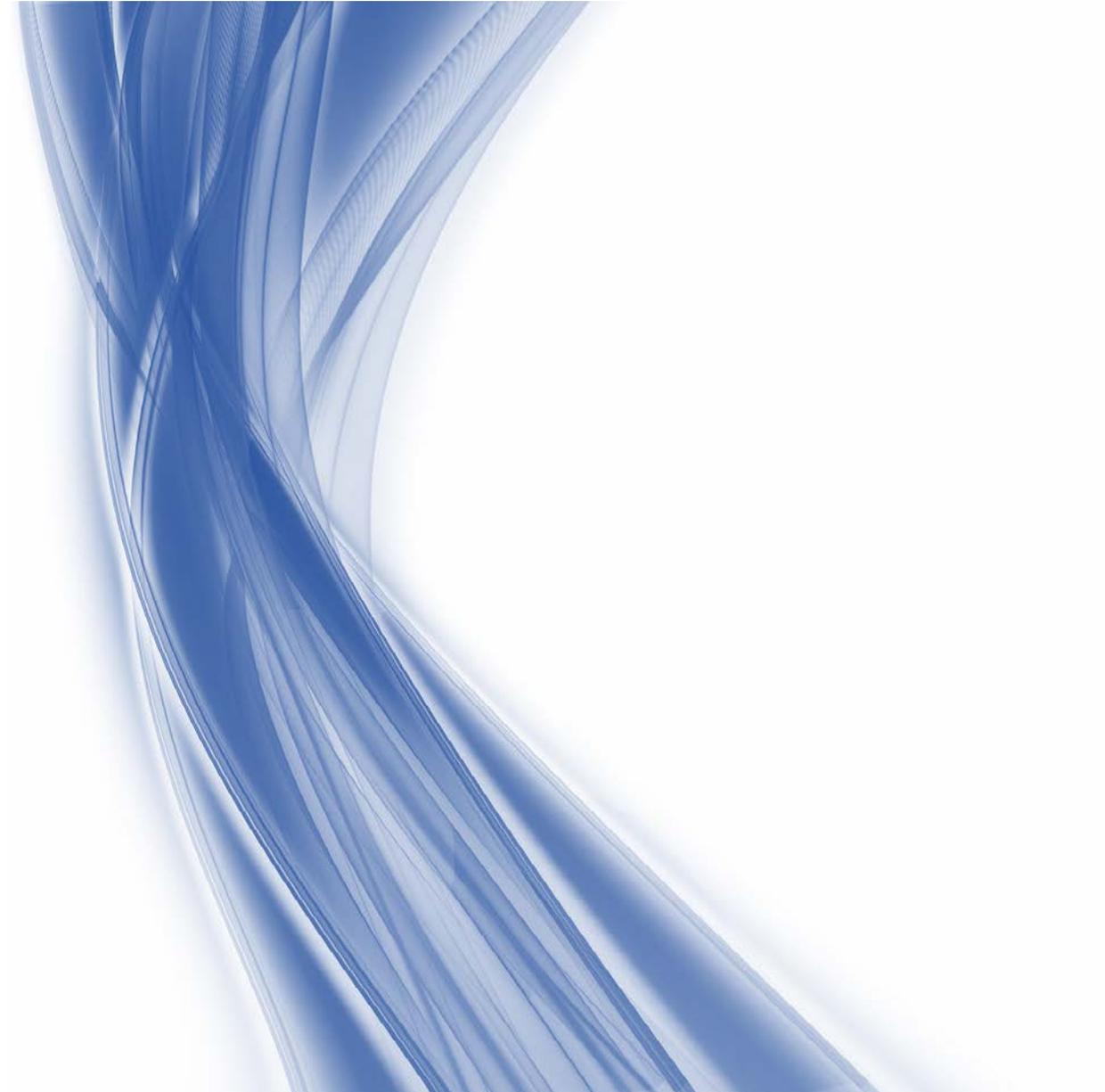


Realizing the Promise of Administrative Data for Enhancing the American Community Survey

U.S. Census Bureau
American Community Survey Office

November 2018



United States™
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Intentionally Blank

Introduction

American society demands increasing amounts of data at a rapid pace to meet the needs of the changing landscape of America's communities. Declining response rates, along with growing concerns about privacy and confidentiality of data, challenge our ability to collect information using surveys. The U.S. Census Bureau has embraced this challenge and is turning to administrative data to strengthen survey programs by easing the burden on Americans asked to respond to federal surveys and deliver more rich, quality data to inform decisions made by policymakers, businesses, and many others.

The American Community Survey (ACS) is the only source of comparable data for all of America's communities. Since 2005, the ACS has provided information on an annual basis that was previously collected only every 10 years on the decennial census. By collecting this information on an annual basis, the Census Bureau has greatly enhanced its ability to meet the needs of our various data users.

The Census Bureau continually researches ways to improve the quality of the ACS data, ensure the survey operates efficiently and is cost-effective, and to ensure the experience of ACS respondents is pleasant and they are asked only what is necessary. Administrative records show great promise for improving the quality of ACS data, enhancing the data products that are available, and asking less of our respondents. This document charts the course for the path forward to integrate administrative records into the ACS program.

What are administrative records?

Administrative records and third party data refer to micro data records contained in files collected and maintained by administrative (i.e., program) agencies and commercial entities. Government and commercial entities maintain these files for the purpose of administering programs and providing services. Administrative records are distinct from systems of information collected exclusively for statistical purposes, such as those the U.S. Census Bureau produces under the authority of Title 13 of the United States Code (U.S.C.). The Census Bureau uses, and seeks to use, administrative records developed by federal agencies, tribal, state, and local governments as well as data from commercial entities.

ACS Program Priorities



The Vision

The Census Bureau is considering several approaches to use administrative records to enhance the ACS program:

- **Reduce the burden on our respondents** by using information already available to the federal government instead of asking questions.
- **Improve data quality** by drawing upon administrative records for data editing and imputation, rather than statistical approaches to assigning values.
- **Create blended data products**, including merging administrative data on topics not currently asked about on the ACS to provide even more detailed information about each of America’s communities and creating complimentary products for small geographies to improve the quality of community-level estimates such as median income.

Using Administrative Records is Mandated by Title 13 of the U.S. Code:

To the maximum extent possible and consistent with the kind, timeliness, quality and scope of the statistics required, the Secretary shall acquire and use information available from any source referred to in subsection (a) or (b) of this section instead of conducting direct inquiries.

Past Experience

The Census Bureau has a long history of using administrative records to provide quality information about the U.S. population and economy. These data have been used for decades to produce population [estimates](#) and [projections](#). The [Longitudinal Employer-Household Dynamics](#) program has been integrating existing data from state-supplied administrative records on workers and employers with existing censuses, surveys, and other administrative records to create a longitudinal data system on U.S. employment since 2000. The Small Area Income and Poverty Estimates ([SAIPE](#)) Program uses administrative records to provide annual estimates of income and poverty statistics for all school districts, counties, and states. The Small Area Health Insurance Estimates ([SAHIE](#)) Program also uses administrative records to develop model-based estimates of health insurance coverage for counties and states. The [2020 Census](#) will use information that people have provided to improve accuracy and reduce the need for census takers to knock on doors. Before the Census Bureau can employ administrative records, each administrative record resource must be examined to evaluate the value and feasibility of relying on it as a substitute for primary survey data.

Guiding Principles

The Census Bureau has established a set of guiding principles to determine what administrative sources are appropriate for use on the ACS. These principles are used to evaluate each potential source to determine the suitability of using the administrative source either in place of asking a question on the survey or for editing and imputation of survey data.

Authorization	Do we have a formal agreement (e.g., contract or interagency agreement) to obtain and use an administrative data sources?
Availability	Are the data available for every year?
Conceptual Alignment	Do the administrative data correspond to the concept the ACS currently intends to measure?
Coverage	How comprehensive is the coverage of the administrative data with respect to geographies and population subgroups?

Data Source	Do the administrative data come from a trusted and respected source, above reproach and conflict free?
Disclosure Avoidance	Does use of the administrative data preclude the Census Bureau from ensuring disclosure avoidance of personally identifiable information?
Impacts on Estimates	To what extent does the administrative data source diverge from survey response (direct impact)? Do the differences carry over to other variables, for example, through editing and imputation (indirect impact)?
Intended Use	How will the administrative data be used (e.g., editing and imputation, substitution, blended data product)?
Population Universe	Are the administrative data intended for use to measure something for the total U.S. population or a population subgroup (e.g., condo owners)?
Quality	What is sufficient data quality for the published estimates? Do the administrative data meet these quality requirements?
Reliability	Are the administrative data available and consistent over time?
Temporal Alignment	Do the administrative data correspond to the time period referenced in the ACS?

Evaluating Administrative Sources

What sources are available?

The Census Bureau currently houses administrative records from federal, state, and third-party sources relevant to 23 ACS content areas (see Table 1 on page 9). Reliability, coverage, and quality of records vary by source.

Federal sources of administrative records used at the Census Bureau include the Internal Revenue Service (IRS), Social Security Administration (SSA), Department of Housing and Urban Development (HUD), U.S. Postal Service (USPS), Selective Service System (SSS), Indian Health System (IHS), and the Centers for Medicaid and Medicare Services. Each provides demographic, social, economic, and/or housing data on individuals, though none of these sources cover the entire U.S. population. Generally speaking, these federal administrative records are appropriate for usage in the ACS, as they meet several of our guiding principles. In particular, federal data are generally trusted and respected, authorized for use, reliable, of high-quality, and provide excellent coverage of the populations they reach.

The Census Bureau acquires data from states for some of the federally-funded benefits programs they administer, such as the Supplemental Nutrition Assistance Program (SNAP), Temporary Assistance for Needy Families (TANF) program, and the Women, Infants, and Children (WIC) program. State TANF records are relevant to several ACS demographic and income measures, while SNAP records are relevant to ACS Food Stamps questions. Relative to federal data sources, however, state-sourced administrative records may be less appropriate for ACS use. The Census Bureau currently has data from only a handful of states, and some states provide only samples of the individuals participating in their programs. Moreover, the quality of these state-provided administrative records, specifically the documentation and completeness, varies from state to state.

The Census Bureau also contracts with third-party firms, which provide demographic characteristics for population subgroups and housing characteristics for select units. Third-party data on individuals are obtained from sources like telephone books, change of address information for periodicals, and telephone, utility, voter registration, and property tax records. Third-party data on households can also come from deeds, foreclosure, and property tax records. Relative to federally-sourced records, third-party data may prove less appropriate for ACS use because of its reliability (e.g., consistency of measures), future availability, and coverage concerns.

Evaluating Administrative Records

The Census Bureau has evaluated the potential applications of these administrative records for the ACS, focusing primarily on five of the guiding principles:

- Coverage
- Quality
- Conceptual Alignment
- Temporal Alignment
- Impacts on Estimates

Demographic Measures

Administrative records show particular promise for several demographic ACS items. As shown in Table 1, most administrative records contain at least some demographic information on the individuals they cover. Additionally, because of the relatively fixed nature of demographic characteristics like sex and date of birth, the conceptual and temporal alignment of the administrative records and ACS measures is high. But, administrative record coverage of ACS respondents varies from source to source, and prior Census Bureau research shows that administrative records cover younger individuals less well than older individuals. Nevertheless, this research also shows that administrative records may prove particularly useful in ACS production of age and sex items.

The potential applications of administrative records race and ethnicity data to ACS production are impacted by conceptual misalignment between the administrative records and ACS items, as well as inconsistent administrative record coverage of racial and ethnic groups. While the Census Bureau treats race and ethnicity in an additive fashion (such that an individual could identify as both Black Alone *and* Hispanic, for example), it is common for administrative records sources to treat race and ethnicity as mutually exclusive (such that the same individual is either Black Alone or Hispanic, but not both). Prior research shows that White alone, Black alone, Asian alone, and non-Hispanic individuals are better covered and exhibit higher rates of administrative records-ACS response agreement than other racial and ethnic groups. One promising opportunity here is the pooling of multiple ARs to harmonize and catalog an individual's race and ethnicity responses across sources and over time, increasing potential utility for ACS use.

Social and Economic Measures

Relative to demographic items like sex and date of birth, a person's social and economic characteristics are more subject to change over time. This fact makes the temporal and conceptual alignment of administrative records and ACS social and economic measures particularly important.

Preliminary Census Bureau research on the use of IRS records for ACS income measures is promising. The utility of IRS records depends largely on the degree to which the IRS tax year coincides with an ACS respondent's period of reference in the survey. Despite temporal alignment issues, IRS administrative records provide substantial coverage for the *incidence* of income from various sources (e.g., wages, self-employment

income, dividend/interest income) included on the ACS, while wage and salary income *amounts* reported in IRS records show only small discrepancies with ACS responses.

IRS records may also be applicable to the ACS Residence One Year Ago (ROYA) question, which asks respondents whether or not they live in the same house they did one year ago. Prior Census Bureau research notes, however, that issues with respect to conceptual and temporal misalignment limit ACS applications to those respondents collecting some sort of income in the U.S. for two consecutive years, who respond to the ACS survey at approximately the same time they file taxes. Nevertheless, IRS records may be particularly useful for ACS ROYA responses that would have otherwise been edited or imputed.

Housing Measures

As shown in Table 1, third party administrative records contain information applicable to several ACS housing characteristic items, such as housing costs, home value, and tenure status. However, the future availability and reliability of third party records is less certain than that of federal and state records. Prior Census Bureau research finds that third party coverage of ACS households is low for measures of heating fuel, plumbing/kitchen facilities, and some mortgage measures. Additionally, preliminary research has found that some measures, such as tenure status, found in third party records do not provide the requisite level of quality for usage in ACS. While these factors limit the applicability of third party records for the direct replacement of ACS items (see the “Case Study” section of this report), research suggests that third party records may be useful in the editing and imputation of some housing measures.

Case Study: Replacing Housing Characteristics Data on the 2015 ACS

In support of the Census Bureau’s goal to replace or supplement survey data to improve survey response and reduce respondent burden, we tested replacing 2015 ACS response data with administrative records for year structure built, acreage, property value, and real estate tax.¹

Administrative Data Source

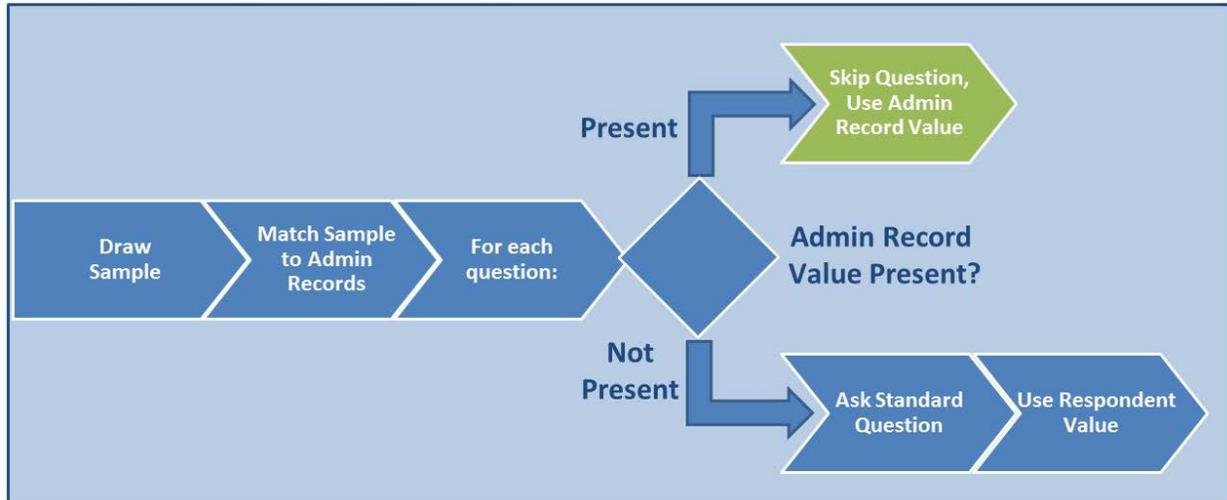
This study uses administrative data from county and municipal tax records obtained from CoreLogic, a commercial vendor.

Adaptive Strategy for Integrating Administrative Data

The goal of a direct-replacement approach for integrating administrative data is to not ask a question if the administrative source provides the information needed. The adaptive design for incorporating administrative data is presented in the figure below. Once administrative data are matched to the survey sample, the records are reviewed to determine if there is an administrative record present. When the record is present, the survey instrument will skip the question (i.e., not ask the respondent) and use the administrative record value. If the record is not present for the sampled address, the survey instrument will ask the standard question and use the respondent value.

¹ Detailed results from this study, when made publically available, will be published at:
<https://www.census.gov/programs-surveys/acs/library/publications-and-working-papers/research-and-evaluation.html>.

Adaptive Design for Integrating Administrative Records in ACS Data Collection



The ACS is collected by self-response via internet or paper questionnaire, and through a computer-assisted personal interview during our nonresponse follow up process. The internet and computer-assisted interview modes use automated data collection instruments that can be programmed to skip questions, making them ideal for incorporating administrative data when available for a specified address. The number of items available in the administrative data for a housing unit varies, which would require multiple versions of the paper questionnaire. This would be difficult and costly to implement and manage. For this reason, the housing study only replaced responses values obtained by internet or computer-assisted interviewing. If the question was left blank on the paper questionnaire, the missing value was filled in using administrative data when available.

Research Findings

We assessed the impact of replacing survey responses with administrative data by:

- Comparing simulated estimates (using administrative records) with published estimates.
- Gauging the reduction in the number of cases that required allocation of values due to missing data.
- Evaluating the potential reduction in burden to respondents by estimating the number of respondents who would not have needed to be asked the question if administrative data were available.

How did the simulated estimates using administrative data differ from our published 2015 estimates?

- There were statistically significant differences for a large proportion of summary metrics for all four items across geographic levels (nation, states, counties, and places).
- At the national level, all but one of the 15 key measures studied were statistically different.
- Many of the simulated estimates had lower measures of property value and real estate taxes than the 2015 ACS estimates.
- The simulated estimates were generally lower than the published estimates for the acreage item.
- Compared to the published estimates, the simulated estimates appeared to have a larger number of housing units in the older year built categories and a smaller number of housing units in the more recently year built categories.

How did using administrative data effect assignment of values due to missing data?

- Using administrative data reduced allocation rates (i.e., assignment of values due to missing data) significantly, ranging from 2.3 percentage points lower to 12.4 percentage points lower.

What potential do administrative data show for reducing the burden on respondents to answer these survey questions?

- We were able to successfully replace responses to the four survey questions in this study for between 37.5 and 53.5 percent of ACS responding housing units.

Using administrative data clearly improves item allocation rates and reduces respondent burden. However, many of the simulated estimates were significantly different from the published estimates. In addition, there were geographic disparities in coverage of the administrative data. Not having complete coverage of administrative data for all geographic areas and housing types means that data for some areas would contain mostly ACS response data, others mostly administrative data, and others with varying combinations of the two sources. Differential coverage and difference between the estimates derived from administrative data versus self-response data are of particular concern, as these issues may compromise the ability of the ACS to represent all areas and housing units as equally as possible.

The results from this study indicate that a direct-replacement method is not appropriate for the ACS, but other methods for incorporating administrative data in place of survey response such as modeling or data harmonization may prove more successful. Improvements in allocation rates suggest that administrative data should be considered for use in our editing and imputation procedures.

Lessons Learned

We learned a great deal from this research and will apply those lessons to our administrative records research moving forward.

1. **We have the ability to bring the administrative data into the production environment.** This simulation successfully demonstrated that we are able to incorporate the administrative data into our production systems, running all required programs and procedures to produce the tables published each year.
2. **Direct replacement alone is not appropriate for integrating administrative data on these selected housing items into the ACS.** This method does not meet several of the guiding principles. The data diverge from survey response and impact other survey items (Impacts on Estimates Principle). They do not offer complete coverage (Coverage Principle), and differ by concept and reference period (Conceptual and Temporal Alignment Principles). Modeling or data harmonization approaches should be considered to ensure estimates are comparable across geographies and better align with the guiding principles.
3. **Administrative data have the potential to improve our editing and imputation methods by using what we know about a housing unit to inform the assignment of a value.** This could benefit the ACS program and satisfies the guiding principle, Intended Use.
4. The data for this study come from an outside vendor. **There are risks associated with using an outside vendor**, such as another vendor being awarded the contract at a future date or the current vendor going out of business. In addition, we need to know if this vendor provides the most comprehensive administrative data or if there are other vendors that do a better job. There will be a need to keep up-to-date on this over time. The guiding principles must be carefully considered when choosing a vendor. It is crucial that the Data Source, Reliability, and Quality principles are fully met.

5. **There were some issues with matching the administrative data to the survey data.** This appears to be concentrated among certain types of housing units, such as multi-unit structures and trailer parks. We must continue work to improve our ability to link data sources and represent all areas of interest (Coverage and Population Universe Principles).
6. **Using administrative data impacts the entire survey life cycle, including the editing and imputation of other items** (e.g., the housing items used in this study are used for editing and imputing income). We must follow our standard practice of thoroughly testing any change to our production system to ensure the Census Bureau's high quality standards are met and that the Quality and Impacts on Estimates Principles are considered in the decision making process.

The Path Forward

Providing a positive experience, minimizing burden for survey respondents, and increasing data quality are clearly and demonstrably a central focus of the ACS program. There is no one-size-fits-all approach for accomplishing this, and the Census Bureau remains agile in researching new ways to create a positive customer experience and improve data quality.

The field of survey research is shifting and we must do what we can to leverage other data sources to enhance, supplement, or in some cases replace what we gain from surveys. Leveraging existing data sources through linked approaches will be an important component of demographic research in the coming years.

Administrative sources show great promise and the Census Bureau is engaging the use of administrative records at all stages of the survey life cycle. We have established a robust administrative records research agenda, seeking to implement administrative data into ACS production as soon as possible, while continuing to adhere to our high quality standards and requirements to ensure we continue to meet the needs of our data users. To meet this goal, we have identified several next steps to build forward momentum:

- **Update and augment existing data sharing agreements** to facilitate the use of administrative records in ACS production.
- **Pursue additional data sharing agreements** to broaden the range of ACS items addressed in administrative records. For example, acquiring administrative records that cover ACS content areas such as disability, employment, school enrollment, and educational attainment may expand the scope of administrative records use in ACS production.
- **Cultivate and expand the acquisition of state administrative records** to increase coverage of ACS respondents, as well as AR-ACS agreement rates.
- Continue research to **improve linkage methods** and **assess the quality of administrative data**.
- Consider how administrative records from various sources could be used to **create harmonized measures of individual and household characteristics**. This may increase coverage of ACS respondents, as well as AR-ACS agreement rates.
- Implement administrative records in the **ACS editing and imputation procedures**.
- Explore more deeply options to **replace or substitute survey responses with administrative data**, with an emphasis on modeling approaches to ensure consistent representation across geographies and population groups, along with rigorous evaluation of impacts to survey estimates.

Table 1: An Evaluation of Administrative Records for American Community Survey Subject Areas

ACS Subject Area	AR Available?	Multiple ARs Available?	Available for Research?	Geographic Level	Observational Unit	ACS Match Rate [†]	Response Agreement Rate [‡]
Demographics							
Age	Yes	Yes	Yes	National	Individual	90-95%	93-97%
Hispanic Origin	Yes	Yes	Yes	National	Individual	37-60%	77-98%
Race	Yes	Yes	Yes	National	Individual	10-72%	17-99%
Relationship	Yes	Yes	Yes	National	Individual	Unknown	Unknown
Sex	Yes	Yes	Yes	National	Individual	93%	99%
Social Characteristics							
Ancestry	Yes	Yes	Yes	National	Individual	Unknown	Unknown
Citizenship	Yes	No	Yes	National	Individual	Unknown	Unknown
Disability	No						
Educational Attainment	No						
Fertility	Yes	No	Yes	National	Individual	Unknown	Unknown
Grandparents	No						
Language	No						
Marital Status	No						
Migration	Yes	Yes	Yes	National	Individual, Household	78-98%	82-85%
School Enrollment	No						
Veterans	Yes	No	Yes	National	Individual	Unknown	Unknown

[†] When multiple administrative record sources are available, the ACS Match Rate for the source with the highest match rates is shown. Rate ranges indicate variation in ACS Match Rates for different subject area questions and/or population subgroups.

[‡] When multiple administrative record sources are available, the Response Agreement Rate for the source with the highest ACS Match Rates is shown. Rate ranges indicate variation in ACS Match Rates for different subject area questions and/or population subgroups.

Table 1, Continued: An Evaluation of Administrative Records for American Community Survey Subject Areas

ACS Subject Area	AR Available?	Multiple ARs Available?	Available for Research?	Geographic Level	Observational Unit	ACS Match Rate [†]	Response Agreement Rate [‡]
Economic Characteristics							
Class of Worker	No						
Commuting	No						
Employment Status	Yes	No	No	National	Individual	Unknown	Unknown
Food Stamps (SNAP)	Yes	Yes	Yes	State	Individual	Unknown	Unknown
Health Insurance	Yes	Yes	Yes	National	Individual	Unknown	Unknown
Hours/Week, Weeks/Year	Yes	No	No	National	Individual	Unknown	Unknown
Income	Yes	Yes	Yes	National	Individual, Household	88-98%	87%
Industry & Occupation	Yes	No	Yes	National	Individual, Household	Unknown	Unknown
Housing Characteristics							
Computer & Internet Usage	No						
Costs (Mortgage, Taxes, Insurance)	Yes	No	Yes	National	Housing Unit	4-77%	13-87%
Heating Fuel	Yes	No	Yes	National	Housing Unit	14%	81%
Occupancy	No						
Plumbing/Kitchen Facilities	Yes	No	Yes	National	Housing Unit	14%	95%
Property Value	Yes	No	Yes	National	Housing Unit	99%	29%
Structure	Yes	Yes	Yes	National	Individual, Housing Unit	Unknown	Unknown
Tenure (Own/Rent)	Yes	No	Yes	National	Housing Unit	12%	62%
Utilities	No						
Vehicles	No						
Year Built/Year Moved In	Yes	Yes	Yes	National	Housing Unit	59-75%	76-78%

[†] When multiple administrative record sources are available, the ACS Match Rate for the source with the highest match rates is shown. Rate ranges indicate variation in ACS Match Rates for different subject area questions and population subgroups.

[‡] When multiple administrative record sources are available, the Response Agreement Rate for the source with the highest ACS Match Rates is shown. Rate ranges indicate variation in ACS Match Rates for different subject area questions and population subgroups.