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

November 15, 2023

2023 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES #ACS23-RER-16

MEMORANDUM FOR ACS Research and Evaluation Workgroup

From: Donna M. Daily
Chief, American Community Survey Office

DONNA DAILY Digitally signed by DONNA DAILY Date: 2023.11.30 15:41:02 -05'00'

Lucia Foster
Chief Economist, Center for Economic Studies

LUCIA FOSTER Digitally signed by LUCIA FOSTER Date: 2023.11.29 12:47:52 -05'00'

David Waddington
Chief, Social, Economic & Housing Statistics Division

DAVID WADDINGTON Digitally signed by DAVID WADDINGTON Date: 2023.12.11 18:15:52 -05'00'

Prepared by: Sandra Clark and Dorothy Barth
American Community Survey Office

Ariel J. Binder, Adji Fatou Diagne,
Nikolas Pharris-Ciurej, and John Voorheis
Center for Economic Studies

Evan Brassell and Christopher Mazur
Social, Economic, and Housing Statistics Division

Subject: Acreage Administrative Records Data Research for the American Community Survey

Attached is the American Community Survey (ACS) Research and Evaluation Report, "Acreage Administrative Records Data Research for the American Community Survey." This report evaluates available administrative records data for acreage and outlines a method to use the data in ACS production. Using that method, we simulated replacing survey data with administrative records data and compared the results to the current production method that strictly uses survey responses. All data in this report have been approved for publication by the Disclosure Review Board with the Approval Number CBDRB-FY23-CES019-014.

If you have any questions about this report, please contact Sandra Clark at 301-763-5884.

Attachment
cc:
acso.re.workgroup.list@census.gov
Katherine Mark ACSO
Jason Lizarraga
Judy Wang
Lucia Foster CES
ces.demo.research.list@census.gov
Dave Waddington SEHSD
Jacob Fabina
Caroline Short

DRB Clearance Number: CBDRB-FY23-CES019-014

American Community Survey Research and Evaluation Program

November 15, 2023

Acreage Administrative Records Data Research for the American Community Survey

FINAL REPORT

Sandra Clark and Dorothy Barth
American Community Survey Office

Ariel J. Binder, Adji Fatou Diagne, Nikolas Pharris-Ciurej, and John Voorheis
Center for Economic Studies

Evan Brassell and Christopher Mazur
Social, Economic, and Housing Statistics Division

This page is intentionally blank.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

EXECUTIVE SUMMARY

The U.S. Census Bureau has a long history of using administrative records (AR) data to provide quality information about the U.S. population and economy. These data have been used for decades to produce population estimates and projections. With decreasing survey response rates and increasing needs for data, programs are researching ways to tap into data from AR sources. The Census Bureau is actively looking for ways to use AR data to enhance the American Community Survey (ACS) and reduce the burden placed on our respondents, improve data quality, and create blended data products to meet data user needs.

One of the ACS questions recently evaluated for AR data usage is the question about acreage, or lot size, for a single-family house or mobile home. We evaluated acreage data from the vendor, Black Knight, Inc., to determine if it was acceptable to use in place of ACS survey data. Our research shows that the vendor data covers about 87 percent of the full ACS sample. Also, when comparing the vendor data to ACS responses for acreage, the vendor data matched 90 percent of the time. The high coverage and match rates are indicators that the vendor data are acceptable to use in place of ACS responses for the acreage question.

After determining acceptable use, we produced a plan for the best way to use the data for the ACS. If acreage data are available for Computer-Assisted Personal Interview (CAPI) and internet responses, we will skip the survey question and use AR data in its place. For now, the acreage question will still be asked on the paper questionnaire and via Telephone Questionnaire Assistance (TQA), but we will replace the survey responses with AR data in post data-collection processing when it is available.

Using this method, we simulated what we may expect to see in production to compare the AR data-use method with the current production method, which only uses survey responses. For each category of acreage (1 = less than an acre, 2 = 1 to 9.9 acres, 3 = 10 or more acres) we compared simulation vs. production estimates by demographic characteristics of the respondent (*age, sex, Hispanic origin,* and *race*) and household characteristics (*occupancy status, type of building, year built,* and *tenure*). We compared the estimates nationally and for each state. For all comparisons, we conducted two-tailed t-tests to detect statistically significant differences at the α=0.1 level.

In the simulation, certain trends appeared across a variety of demographic, geographic, and structural characteristics. Including the District of Columbia, nearly half of states (24) had a significant increase in the proportion of units on *less than an acre*. The proportion of units on *10 or more acres* increased for thirty-one states.

This is important to note because one of the main uses for the acreage item is to identify housing units on large lots. The Department of Housing and Urban Development (HUD) and the Bureau of Economic Analysis both subset out these units for their calculations (e.g., HUD's Fair Market Rents). While the trends in aggregate indicate that the simulated data tends to have

more housing units reporting their lot is under an acre, this is mostly driven by the single-family detached structure type. This is by far the largest category of housing units in the country, so any trend with those housing units can mask other impacts.

When looking at mobile homes and single-family attached houses, both had a significant decrease in the percentage of housing units on less than an acre, which runs counter to the overall trend. This corresponds with a significant increase in both the *1 to 9.9 acres* and *10 or more acres* categories for both types of structures (mobile homes and single-family attached houses). In the simulated data, 8% of mobile homes were on ten or more acres, as opposed to the 2019 production ACS where 6.2% of mobile homes were on these large lots. Single-family attached houses also had an increase, from 0.6% to 1.4% of this type of structure now being on ten or more acres.

The above breakdown is also apparent based on tenure categories. Rented households (both with cash rent and those occupied without paid rent) and those owned without a mortgage had increases in the number of units on ten or more acres, but both categories of rented units had much larger changes in their proportions. Units with cash rent went from 1.8% of their units on ten or more acres to 3.4%. Units not paying cash rent went from 10% on large lots to 13.1%. For owned units, the trend mimicked the overall trend, with a shift from units on 1 to 9.9 acres to units on under an acre.

Many housing items have differing item nonresponse rates based on the tenure (owned vs. rented of the unit), and acreage does as well. Owned units had an item nonresponse rate of 3% in the 2019 ACS for acreage, while rented units had a nonresponse rate of 4.8%. Using administrative data can help mitigate any bias that the increased imputation rates for rented units introduce.

While the research points to some possible impacts on ACS estimates, we believe that the benefits of using administrative data in place of survey responses for acreage outweigh any of the possible downfalls. Incorporating this data will help reduce respondent burden and potentially increase data quality for a vast majority of housing units. This change will also bring the ACS in line with the other Census survey that also collects this information, the American Housing Survey, which has been using a similar methodology for nearly a decade. We recommend implementing administrative records data for acreage into full ACS production according to the methodology outlined in this report. We will continuously monitor the use of administrative acreage data and we may update the methodology when necessary or possible.

1. BACKGROUND

The American Community Survey (ACS), conducted by the U.S. Census Bureau, is an ongoing nationwide survey with twelve monthly panels. The data collection process for each monthly panel spans a three-month period. The first two months are dedicated to self-response by internet, mail, and phone (via a Telephone Questionnaire Assistance line). While self-responses are still accepted in the third month, data collection activities are focused on Computer-Assisted Personal Interviews (CAPI) via a phone call or a personal visit from a Census Bureau Field Representative to the housing units in sample.

With decreasing survey response rates, programs are researching ways to tap into data from administrative record (AR) sources. The Census Bureau has a long history of using AR data 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 Census Bureau is actively looking for ways to use AR data to enhance the ACS to reduce the burden placed on our respondents, improve data quality, and create blended data products to meet data user needs.

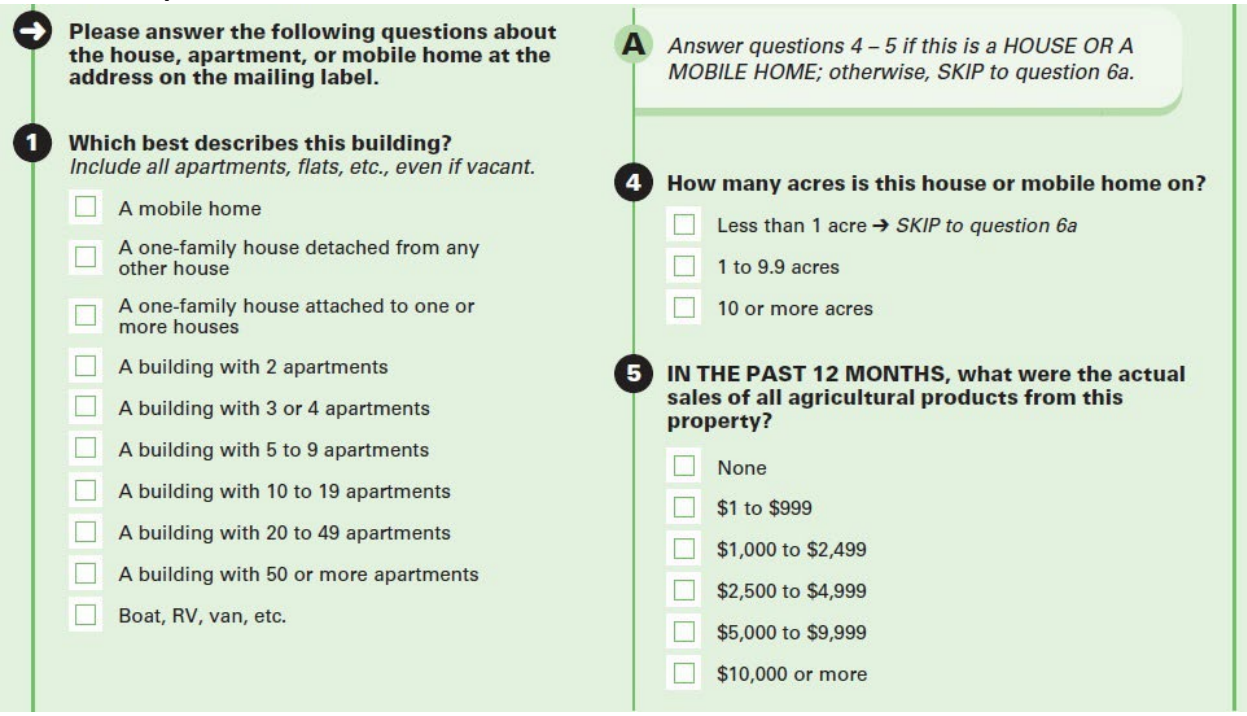
One of the ACS questions recently evaluated for AR data usage is the question about acreage, or lot size, for a single-family house or mobile home. After evaluating acreage data from a third-party vendor, we determined that the data are acceptable to use in place of survey data. If acreage data are available for CAPI and internet responses, we will skip the survey question and use AR data in its place. For now, the question will still be asked on the paper questionnaire and via Telephone Questionnaire Assistance (TQA), but during post-processing edits, we will replace acreage survey responses with AR data when possible.

Using this method, we simulated what we may expect to see in production to compare the AR data-use method with the current production method, which only uses survey responses. This document contains some background information about the data, the results of acreage administrative records data analysis, and the results of the simulation.

1.1 ACS Acreage Question

The ACS contains three sections: a basic demographic section about persons living in the household, a detailed housing section, and a detailed section about persons living in the household. The first question in the ACS housing section is about the type of building at the sample address. If it is a mobile home or single-family house, then the question about acreage is asked. If the mobile home or house is occupied and is on one or more acres, then the question about agricultural sales from the property is asked. Because the questions are related to one another, the analyses and results presented in this report account for the interaction of the three questions. See Figure 1 for an image of the questions on the paper questionnaire.

Figure 1. Image of the Type of Building, Acreage, and Agricultural Sales Questions on the 2023 ACS Paper Questionnaire



The statistics created from the question about acreage are used by multiple government agencies including the Department of Agriculture, Department of Commerce, Bureau of Economic Analysis, Department of Housing and Urban Development (HUD), and the Environmental Protection Agency. These agencies use the data for a variety of purposes, like identifying the rental distribution of housing units to determine Fair Market Rent, creating Gross Domestic Product estimates, and researching pollution and development patterns. Community planners can also use this information to better understand the local housing markets, identify possible zoning changes, and reduce tax revenue losses from vacant or abandoned properties. All these needs make the data gathered by the ACS important for the federal government and the public.

1.2 Administrative Records on Acreage

1.2.1 Overview of Acreage Data

County assessors collect and maintain information about local housing stocks for property taxation and other administrative purposes (e.g., zoning, deeding, and permitting). These records are publicly available and contain information related to some of the information that is elicited on the ACS and other household surveys. For example, in addition to lot acreage, these records contain information on when the structure was built, property taxes paid on the structure, square footage, number of bedrooms, and some history of previous transactions, among many other items (Molfino et al., 2017; Weinberg, 2015). In most cases, this information

is sourced directly from property tax assessment records. Hence, throughout this report, we refer to "property tax records" and "administrative records" interchangeably.

A robust industry of third-party vendors has emerged in recent years that aggregate and standardize property tax and related administrative records across thousands of local jurisdictions, creating national-level data products available for purchase.[1] For several years, the Census Bureau has maintained a contract with a third-party vendor, and research teams have investigated various possibilities of improving household survey data collection or production with these administrative-record data products (Brummet, 2015; Ruggles, 2015; Clark et al., 2018; Dillon, 2019; U.S. Census Bureau, 2020).[2] The American Housing Survey (AHS) has used administrative records in its imputation models for the acreage and year-built items since 2015 (Molfino, 2021a and 2021b). Our report draws on this and our own research to design and test an adaptive specification that will remove the acreage question, in favor of direct replacement with administrative records, for most ACS respondents.

The Census Bureau's current vendor of property tax records is Black Knight, Incorporated (BK). While the facts and considerations reported below relate to this vendor, it is plausible that they generalize to other vendors as well, given that the underlying source of information (county assessor records) is the same. Moreover, prior analysis of property tax records from two anonymous vendors found that the two sets of records had extremely similar rates of alignment with the AHS across various survey topics, including acreage (Binder et al., 2022).

1.2.2 Potential Misalignment with Household Survey Data

Prior research led us to conclude that the acreage item would be a prime candidate for replacement or extensive supplementation with property tax records (Dillon, 2019). Nonetheless, we note several sources of potential misalignment between property tax records and household survey information and discuss upfront how these misalignments might affect our adaptive design.

First, there is a degree of *conceptual misalignment* between the two information sources. Household surveys such as the ACS collect information about *housing units*. Administrative records contain information about *parcels* of land that are subject to property taxes. Housing units and parcels are not generally the same thing—for example, an apartment building containing multiple units is usually taxed as one parcel. In such a case, the administrative record

[1] The records we will be considering in this research exist for all 50 U.S. states and the District of Columbia, but do not exist for Puerto Rico and other U.S. territories.

[2] The Census Bureau acquires these administrative records through a competitive acquisition process that solicits proposals from various data vendors. The typical contract awarded to the winning proposal is for a term of five years, after which the Census Bureau solicits a new set of proposals for subsequent acquisitions. The first contract was awarded in 2014, and a subsequent one was awarded in 2019. The next contract is slated to be awarded in 2024.

applies to the whole parcel rather than specific units on the parcel. Moreover, there are cases in which the legal address of the parcel is a business office where property tax bills are mailed. In these cases, the difference in addresses between the housing unit(s) located on the parcel and the parcel itself complicates or prohibits the linkage of parcel-level information to housing units. Conceptual misalignment presents less of a problem for acreage than for other household survey items since the acreage universe consists only of single-family housing units, where parcel and housing-unit boundaries (and addresses) tend to be identical. However, there are two exceptions regarding mobile homes or single-family attached houses, commonly referred to as "townhouses" or "row houses." Just as for apartments, mobile homes and townhouses sometimes sit on a bigger parcel that has a separate legal address, or where the administrative acreage value applies to the entire development as opposed to the boundaries ascribed to an individual unit by its owner.

Second, the inevitable lag between property assessments and delivery of the vendor file to the Census Bureau creates a degree of *temporal misalignment*. For example, the Census Bureau received the 2022 vintage of BK data in September 2022. However, most of the property assessment dates contained in this vintage were from 2021. Therefore, if we had decided to use administrative records to collect acreage information for the 2023 ACS, most information would be two years out of date. Fortunately, this issue is also less of a problem for acreage than it might be for other survey items since lot acreage is essentially an unchanging feature of a housing unit.[3]

Third, even conditional on perfect conceptual and temporal alignment, there may be *value misalignment* between administrative and respondent-elicited acreage. One source of value misalignment could be caused by the fact that ACS responses are collected as three categorical bins whereas administrative data are typically reported as numeric values of acreage that will to be converted to the ACS categories. Also, subject matter experts suggest that some respondents, particularly renters, may have little offhand knowledge of their lot acreage; while others may have an accurate perception but may round to the nearest whole number—particularly if they live on a slightly below one acre lot. The rounding issue is especially salient because the ACS acreage categories distinguish between lots that are less than one acre versus one acre or greater. Value misalignment can also occur when there are errors or out-of-date information in the administrative data. This data is collected from thousands of jurisdictions, and while it is cleaned and harmonized by the vendors, differences in the data can remain. Also, the ways areas collect this information, and the amount they collect, is not standard. This can result in differences between the collected survey data and the vendor data.

[3] One implication of this lag, that potentially affects acreage just as much as other survey items, is that property tax records will not reflect new construction completed during the year prior to the survey year. Census Bureau estimates on new residential construction, however, suggest that this is figure is a very small share of the total housing stock. (See Binder et al. 2022 for further discussion.)

When replacing a less accurate with a potentially more accurate source of information, removing rounding bias or inaccuracies in the data collected from the jurisdictions may have downstream effects on other information that is dependent on acreage, such as reporting for agricultural sales and the exclusion of large parcels in HUD's computation of Fair Market Rents.

1.2.3 Administrative Records Linkage to Sampled Housing Units

Black Knight provides two independent sources of parcel locations and acreages that can be linked to the housing units sampled by the ACS.

First, the property assessment files contain the parcel's physical address and the assessed acreage as of the most recent assessment. When BK data are first delivered to the Census Bureau, each parcel is assigned an anonymized housing-unit identifier from the Census Master Address File (MAFID) whose address text most closely matches that of the parcel's physical address, using a standard enterprise linkage algorithm. If no suitable housing-unit address is found, the parcel is not assigned any MAFID. For example, 62.7% of all parcels in the 2019 BK assessment file were assigned a valid MAFID (Binder et al., 2022). We should not expect a 100% match rate because many taxed parcels have non-residential land uses.

Second, BK provides a set of county-level parcel boundary files, along with the calculated area (in acres) of each parcel based on these contiguous geographic polygons. Nearly all MAFIDs in the MAF contain a set of geographic coordinates; it is therefore possible to link almost all parcels to an associated housing unit via a geospatial join routine. The geospatial intersection identifies the polygon associated with a given MAFID's latitude and longitude. Note that in cases in which there are multiple units on a given property, this intersection will map multiple MAFIDs to an individual parcel, unlike the address-based matching routine, which will assign a single MAFID to a parcel.

1.3 Acreage Administrative Records Data Use on the ACS

Three competing goals informed our design for using acreage administrative records data on the ACS. First, we aimed to reduce respondent burden, as well as reliance on potentially inaccurate responses, by maximizing the usage of administrative records. Second, we aimed to avoid substantially different acreage estimates than the current production method, which could cause "series jumps" in downstream products that rely on acreage information (e.g., agricultural sales estimates and HUD's Fair Market Rents calculations). Third, we aimed to develop a parsimonious file build that could be easily replicated in actual data collection situations. We used these administrative records to design a composite acreage file, which we then tested for possible usage in an adaptive design. We discuss how these goals informed several decision points, and the subsequent composite file specification, below.

1.3.1 Panel Structure of the Administrative Records

A given delivery of the BK assessment file is intended to be comprehensive; parcels not assessed in the vintage year tend to have information allocated from the most recent assessment date. However, we noticed a small amount of slippage between assessment years, likely not just due to new construction or demolition. This suggested to us that including multiple years of BK records in a composite file instead of one year could further reduce respondent burden. Due to the small degree of coverage gains from using two years rather than one, and the goal of developing a parsimonious and straightforward specification, we settled on two years as the optimal number (rather than three or more). We also note that an adaptive design that considers only one year of BK information would still be viable, should data availability issues arise in future.

1.3.2 Address-Based versus Geospatial Information

Given that BK delivers two independent sources of acreage information, we explored how best to combine the two sources into a composite acreage file. The advantage of the geospatial data, relative to the address-based information, is that it suffers less from the conceptual alignment issue. That is, because it does not rely on (potentially incongruent) addresses of housing units versus parcels, but instead on geographic coordinates, the geospatial linking method produced a substantially higher coverage rate of the ACS universe than did the MAFID-based linking method. However, we found that while the geospatial method produced additional links, these links were not always reliable. Specifically, for single-family attached houses and (especially) for mobile homes we found substantially different acreage distributions between the geospatial data and the address-based data (with the latter estimates also being closer to the actual 2019 survey responses). In several subsamples we considered the geospatial linkages produced implausibly large acreage estimates for these building types, suggesting that the parcel boundaries encompassed entire townhouse developments or mobile home parks.

While a more complex design would parse out which counties provide reliable geospatial data, our conservative approach—which also allowed for a more parsimonious file design—was to ignore geospatial information for single-family attached houses and mobile homes. This choice was also driven by the fact that most housing units in the acreage universe are single-family detached houses, where the geospatial data appear no less reliable than the address-based data.

1.3.3 Counties with Poor Data Quality

A small set of counties, often in rural areas with nonstandard housing stock (e.g., many mobile home parks), provided administrative record data that was largely missing or deviated substantially from acreage responses to the 2019 ACS. We attributed these differences to severe conceptual misalignment or possible errors in the administrative records themselves. To screen out these counties that are likely providing data of poor quality, we developed a set of business rules.

The first rule is that a county needed to have a coverage rate of at least 50%. This means that over half of the cases in sample for the 2019 ACS also had to be present on the administrative data files. This rule was put into place to remove the chance of potentially introducing bias into the estimates. Having over half of the data coming from respondents (with the potential of introducing rounding error but limiting conceptual misalignment) and a sizeable subset coming from administrative data, raised concerns that this work would introduce differing errors across cases in a non-random fashion. Furthermore, because of the nature of the third-party data, its availability might differ across certain sub-county geographies (e.g., Census place within a county) or structure types (e.g., mobile homes)

The second rule is that linked historic ACS respondent data needed to match the administrative data 80% of the time for a county's data to be included. While there is the potential that the administrative data might be more accurate for these counties than the respondent data, because of issues mentioned earlier like rounding and nonresponse, we decided to hold back these counties until we performed additional research to identify causes of differences. Some of the known issues with the third-party data, namely the conceptual misalignment, also fed into the decision to require a high level of agreement between the two sources.

Of the 3,143 counties and county equivalents, 321 counties were flagged as not meeting the requirements. Fortunately, despite the number of counties flagged, these business rules flagged a small subset of cases; only 2.8% of MAFIDs sampled in the 2019 ACS resided in a flagged county. Most counties provide acreage data of sufficient quality for use in our composite file.

These metrics are based on county-level aggregates for two main reasons. First, counties are critical for sampling, weighting, and internal data review procedures. Second, HUD's Fair Market Rent calculations, a primary use case of the acreage data, are done at the county level. These factors helped lead the research towards the approach of an adaptive design, and using the administrative data when possible, but collecting data from respondents when the administrative data did not meet our quality standards.

1.3.4 Composite File Specification

After a comprehensive review of the administrative acreage data, we used the following methodology to prepare the records, as stated below, to assess use on the ACS. Starting with the list of 2019 ACS MAFIDs and geographic coordinates, we built a composite file of administrative acreage records link to the ACS sample as follows:

- MAFID-merged on assessed acreage information from each of the 2021 and 2019 BK assessment files.[4]
- Geospatially merged on calculated acreage information from each of the 2021 and 2019 BK parcel boundary files.
- Recoded all four acreage measures to match the ACS acreage categories:
  - 1: Less than one acre[5]
  - 2: Greater than or equal to one acre, but less than 10 acres
  - 3: Greater than or equal to 10 acres
- Assigned an administrative-data based acreage variable:
  - Started by using the recoded 2021 assessed acreage value.
  - If missing, used recoded 2019 assessed acreage value.
  - If still missing AND building type was a single-family attached house, used recoded 2021 calculated acreage value from the geospatial merge.
  - If still missing AND building type was a single-family attached house, used recoded 2019 calculated acreage value.
  - Finally, reassigned the administrative-data-based variable to missing if the MAFID was in one of the flagged counties.

2. METHODOLOGY

First, we evaluated the composite file data for its coverage and reliability. For acreage, *coverage* refers to the percentage of houses or mobile homes in a yearly ACS sample that can be linked to the commercial AR acreage data. To justify the cost of AR data, the data must be able to replace survey data for a large majority of households in the ACS sample. *Agreement* refers to how well the AR data agree or match with ACS responses. We assumed the AR data would be the same or higher quality than survey responses, so we performed an analysis to verify the assumption. The research was guided by the questions outlined in Section 2.1.

After evaluating the data, we plan to use adaptive design methodology for using administrative acreage data on the ACS, starting with the 2024 ACS data year. Using this methodology, we

[4] The Census Bureau currently only has a business need for BK geospatial data every other year, so 2020 data were not available for this research.

[5] A small share of property assessments recorded an exact zero value for acreage. These zero values did not appear to correlate with ACS responses, so we interpreted them as recording errors and reassigned them to "missing."

simulated what might happen in production and compared the results to production results using the 2019 ACS data. We evaluated the effect on response burden and the effect on the ACS acreage estimates. Because the agricultural sales question is only asked for addresses with one or more acres that are occupied, we also evaluated the effect of the simulation on agricultural sales. The questions that guided the evaluation are outlined in Section 2.2.

2.1 Coverage and Agreement Provided by the Composite File Data

RQ1. What percentage of households in the 2019 ACS sample matched a non-missing administrative acreage record from the composite file?

This rate is defined as the coverage rate. For the coverage rate, we calculated the rate at which records in the ACS sample had corresponding valid and non-missing administrative record values for acreage.

RQ2. How closely does the composite file's information match the acreage responses we received in the 2019 ACS?

This rate is defined as the agreement rate. For ACS sample records with responses to the acreage question, we calculated the rate at which the responses agreed or matched with their corresponding administrative record acreage values.

RQ3. How do the coverage and agreement rates vary across common demographic characteristics of the respondent (age, sex, race, and Hispanic origin)?

RQ4. How do the coverage and agreement rates vary across housing characteristics (occupancy status, building type, year built, and tenure)?

RQ5. How do the coverage and agreement rates vary by Metropolitan and Micropolitan areas?

RQ6. How do the coverage and agreement rates vary by state?

RQ7. When disagreements occur, does the composite file systematically report larger (or smaller) acreage values than the ACS responses?

We computed a confusion matrix at the national level to understand how the administrative record acreage values and ACS acreage responses disagreed. In prior research, not reported here for the sake of brevity, we performed similar disagreement analyses across different housing characteristics. These disagreements will be implicitly discussed in the results section, which analyzes how acreage estimates might change across many of the subgroups if survey responses were replaced with composite-file data.

2.2 Effect of the Composite File Data on ACS Acreage and Agricultural Sales Estimates

The adaptive design methodology to use administrative records on the ACS is as follows: For addresses in counties that have non-missing acreage data that meet the requirements (at least 50% coverage and 80% agreement), we will use the data from the composite file. For internet and CAPI responses, the acreage question will not be asked. For paper and TQA responses, the acreage question will be asked, but the responses will be replaced with administrative record data. If the acreage value is less than one acre, the agricultural sales question will not be asked. For addresses that have missing values on the composite file, the acreage question (and possibly the agricultural sales question) will be asked.

We evaluated the methodology by using the research questions outlined below.

2.2.1 Effect on ACS Acreage Estimates

RQ8. How might using AR acreage data affect tabular estimates when compared to estimates calculated using the current ACS methodology?

To assess how the use of AR data may affect tabular estimates of the ACS data (published estimates or statistics calculated from the public use file), we calculated distributions of respondent demographics, housing characteristics, and geography for the three acreage categories. We calculated and compared the distributions for the simulations vs. production methodology for the following categories:

a) Demographics of the respondent—*age, sex, race, and Hispanic origin.*
b) Housing characteristics—*occupancy status, building type, year built,* and *tenure*.
c) Geography—State

2.2.2 Effect on ACS Agricultural Sales Estimates

RQ9. How does using AR acreage data affect tabular estimates for Agricultural Sales when compared to estimates calculated using the current ACS methodology?

Because the agricultural sales question (AGS) is asked based on the acreage categories, we evaluated the effect of the simulation on agricultural sales estimates by analyzing the effect on the imputation of the agricultural sales value that resulted because of a change in the acreage category from the administrative records data.

3. ASSUMPTIONS AND LIMITATIONS

As discussed above, several features of property tax records limit our ability to remove the acreage question from the ACS instrument and simply fill in the values with these records. We review the assumptions and possible limitations behind our adaptive design below.

3.1 Assumptions

In constructing the composite file to be used in the adaptive design, we made three sets of assumptions.

- First, we used two vintage years of BK data to construct the file, as this offered us higher coverage rates over one year but is less vulnerable to changes in availability or quality of vendor data than using three or more vintage years.
- Second, we opted to use both geospatial and address-based records to maximize coverage but assumed that the geospatial data were not of sufficient reliability to be used for mobile homes or single-family attached houses. Moreover, we assumed that the address-based records were more reliable in cases where both types of records existed but disagreed about the acreage amount.
- Third, given that administrative records for some counties appear to have large gaps or other quality issues, we had to make assumptions to decide which counties' records were of sufficient quality for inclusion in the composite file. These assumptions are the business rules described above, i.e., we require at least a 50% coverage rate and an 80% conditional agreement rate with 2019 ACS responses.[6]

3.2 Limitations

- Our composite file is limited by the inherent conceptual and temporal misalignment issues described above. Specifically, because the administrative records contain information about legal parcels, which may contain different addresses than housing units, or group multiple housing units together in cases of mobile homes or townhouses, it was impossible to perfectly link each housing unit to an analogous property tax record. This limitation caused both imperfect coverage as well as imperfect agreement, which guided our above assumptions to maximize coverage of the composite file while also preserving reliability and agreement with prior ACS responses.
- Regarding temporal misalignment, one further limitation is worthy of note. Because there is roughly a two- or three-year lag between when an assessment is performed and when the given record would be used in a given ACS production year, coverage rates will likely be especially low for newly constructed houses.

[6] We will analyze the administrative acreage data every five years, or if the data come from a new vendor, to update these business rules, if necessary.

4. RESULTS

4.1 Composite File Coverage and Agreement Results

RQ1. What percentage of households in the 2019 ACS sample matched a non-missing administrative acreage record from the composite file?

84.5% of ACS MAFIDs from the 2019 ACS sample contained a valid administrative acreage record.

RQ2. How closely does the composite file's information match the acreage responses we received in the 2019 ACS?

90.2% of the acreage values on the 2019 ACS sample matched the corresponding acreage values on the composite file. We refer to this as the agreement rate.

RQ3. How do the coverage and agreement rates vary across common demographic characteristics of the respondent (age, sex, race, and Hispanic origin)?

Coverage and agreement rates vary little across demographic subgroups, with several exceptions. (See Table 1.) Young householders, Hispanic or Latino householders, American Indian or Alaska Native householders, and Some Other Race householders have coverage rates at least five percentage points below the national rate. The American Indian or Alaska Native group has a coverage rate twenty percentage points below the national rate. Part of the differences observed is because of the disparity in coverage of the administrative data across states. Some states (for example, Hawaii and Alaska) have higher concentrations of some racial groups, and therefore differences in the availability and quality of the third-party data will impact agreement and coverage rates across these groups. These discrepancies are also likely due to additional factors that are often correlated—for example, American Indian or Alaska Native householders are likelier to live in rural areas or areas with large, subdivided parcels (e.g., mobile home parks) where coverage rates tend to be lower or could be concentrated in states with lower coverage rates.

There is less variation in agreement rates across demographic subgroups. All subgroups, except for the American Indian or Alaska Native group and the Pacific Islander group, had agreement rates within four percentage points of the national rate. Moreover, only the former group had an agreement rate below 80%.

Table 1. Coverage and Agreement Rates: Respondent Demographic Distributions

Respondent Characteristics	Coverage %	Coverage S.E.	Agreement %	Agreement S.E.
Overall Coverage and Agreement	84.5	<0.1	90.2	<0.1
AGE	-	-	-	-
Under 30 years old	78.8	0.1	89.6	0.1
30 to 39 years old	85.6	<0.1	91.5	0.1
40 to 49 years old	87.3	<0.1	91.6	0.1
50 to 59 years old	87.6	<0.1	90.7	0.1
60 to 69 years old	87.2	<0.1	90.6	0.1
70 years old or older	86.3	<0.1	89.5	0.1
SEX	-	-	-	-
Female	85.5	<0.1	90.5	<0.1
Male	87.1	<0.1	90.8	<0.1
HISPANIC ORIGIN	-	-	-	-
Hispanic or Latino	78.2	0.1	93.1	0.1
Not Hispanic or Latino	87.5	<0.1	90.4	<0.1
RACE	-	-	-	-
White alone	87.1	<0.1	90.7	<0.1
Black or African American alone	82.7	0.1	89.6	0.1
American Indian and Alaska Native alone	64.2	0.4	78.9	0.5
Asian alone	91.2	0.1	94.0	0.1
Native Hawaiian and Other Pacific Islander alone	83.9	1.0	84.5	1.0
Some Other Race alone	78.8	0.2	92.6	0.2
Two or More Races	81.8	0.2	89.7	0.2

Sources: U.S. Census Bureau: 2019 American Community Survey; Black Knight, Inc.: 2019 and 2021 Assessment and Parcel Boundary Files; DRB Approval Number: CBDRB-FY23-CES019-014

RQ4. How do the coverage and agreement rates vary across housing characteristics (occupancy status, building type, year built, and tenure)?

Table 2 records coverage and agreement rates by housing characteristics. This table provides evidence of some points suggested in Section 1.2. For example, due to the conceptual misalignment issues, single-family attached houses and (especially) mobile homes have substantially lower coverage rates than single-family detached houses. Additionally, mobile homes have lower agreement rates than permanent single-family houses, although even mobile homes have a fairly high agreement rate of 76.8%. Vacant houses have a substantially lower coverage rate, and a slightly lower agreement rate, than occupied houses. Rented houses have substantially lower coverage rates, and slightly lower agreement rates, than owned houses. Finally, while coverage and agreement rates are quite high regardless of when the structure was built, the most recent year-built group (2018 and 2019) had the lowest coverage rate of any group. We suspect that the coverage of recently built structures will be lower in production, since we had the 2021 file available to us for this comparison to 2019 data. In the 2024 production year, the most recent BK data available to us will likely be the 2023 vintage.

Table 2. Coverage and Agreement Rates: Respondent Housing Characteristic Distributions

Responding Household Characteristics	Coverage %	Coverage S.E.	Agreement %	Agreement S.E.
Overall Coverage and Agreement	84.5	<0.1	90.2	<0.1
OCCUPANCY STATUS	-	-	-	-
Occupied	86.3	<0.1	90.7	<0.1
Vacant	70.1	0.1	85.9	0.1
BUILDING TYPE	-	-	-	-
Mobile home	42.8	0.1	76.8	0.2
Single-family, detached	90.6	<0.1	90.7	<0.1
Single-family, attached	62.7	0.1	92.5	0.1
YEAR BUILT	-	-	-	-
2018 and 2019	80.9	0.4	92.3	0.3
2010 to 2017	82.4	0.1	92.0	0.1
2000 to 2009	84.8	<0.1	91.1	0.1
1970 to 1999	81.3	<0.1	88.8	<0.1
1940 to 1969	88.2	<0.1	91.2	<0.1
1939 or earlier	88.0	<0.1	90.7	0.1
TENURE	-	-	-	-
Owned with mortgage or loan	92.0	<0.1	92.5	<0.1
Owned free and clear	84.0	<0.1	89.2	0.1
Rent with payment	75.0	0.1	88.0	0.1
Occupied without payment of rent	70.2	0.3	79.8	0.3

Sources: U.S. Census Bureau: 2019 American Community Survey; Black Knight, Inc.: 2019 and 2021 Assessment and Parcel Boundary Files; DRB Approval Number: CBDRB-FY23-CES019-014

RQ5. How do the coverage and agreement rates vary by Metropolitan and Micropolitan areas?

Table 3 shows coverage and rates according to whether the housing unit resides in a metropolitan statistical, micropolitan statistical area, or other location. Coverage and agreement rates tend to be highest in metropolitan areas and lowest in areas outside of metro- or micropolitan areas. These findings are also consistent with lower data quality from lower-population and rural county offices or areas with large stocks of mobile homes.

Table 3. Coverage and Agreement Rates by Statistical Areas

Statistical Area	Coverage %	Coverage S.E.	Agreement %	Agreement S.E.
Metropolitan	86.9	<0.1	91.9	<0.1
Micropolitan	78.6	<0.1	83.9	<0.1
Other	69.1	0.1	79.8	0.1

Sources: U.S. Census Bureau: 2019 American Community Survey; Black Knight, Inc.: 2019 and 2021 Assessment and Parcel Boundary Files; DRB Approval Number: CBDRB-FY23-CES019-014

Note: Metropolitan and micropolitan areas are defined by the U.S. Office of Management and Budget.

RQ6. How do the coverage and agreement rates vary across states?

Tables 4 shows coverage and agreement rates across the fifty states plus the District of Columbia.[7] While no state has perfect coverage, most states lie in a narrow (and high) coverage range. For example, Indiana has the highest coverage rate at 92.8%, and thirty-five states have coverage rates between 83% and 93%.[8] The next twelve lowest-coverage states all have coverage rates of at least 65%. The four lowest-coverage states are as follows: Alaska at 63.0%, South Dakota at 62.5%, Louisiana at 38.3%, and West Virginia at 20.1%. In addition to those states containing low population density and relatively large stocks of mobile homes, we noted substantial data quality issues in several counties belonging to these states, especially in Louisiana and West Virginia. One example is that some counties only had properties greater than an acre listed on the administrative data. These quality issues were in line with some of the conceptual misalignment issues mentioned above, but may also stem from omissions of, or errors in, certain records themselves.

We find a somewhat similar pattern of state heterogeneity for agreement rates. The District of Columbia, a metropolitan territory comprised mostly of small parcels, has the highest agreement rate at 97.5%, and thirty-five states have agreement rates between 88% and 98%. The next fifteen lower-agreement states all still have agreement rates of above 80%, and the lowest-agreement state, Mississippi, has a 75.2% agreement rate. These findings provide reassurance that our business rules are successfully flagging counties that provide unreliable data. However, given that only a very small share of housing units are located in poor-quality counties, these results indicate that the majority of county offices throughout the country provide reliable acreage records.

[7] BK does not report data for Puerto Rico, so the coverage rate for that territory is exactly zero and the agreement rate is therefore undefined.

[8] We cannot reject that Indiana and Kansas have the same coverage rate, but Indiana has a statistically significantly higher coverage rate than all other states.

Table 4. Coverage and Agreement Rates by State: Alabama to South Carolina

State	Coverage %	Coverage S.E.	Agreement %	Agreement S.E.
Alabama	69.7	0.3	80.9	0.3
Alaska	63.0	0.6	87.7	0.6
Arizona	86.7	0.2	94.8	0.1
Arkansas	76.2	0.3	82.5	0.3
California	89.7	<0.1	93.5	0.1
Colorado	85.1	0.2	94.5	0.1
Connecticut	88.8	0.3	87.5	0.3
Delaware	88.0	0.4	91.5	0.4
District of Columbia	91.0	0.7	97.5	0.4
Florida	89.7	0.1	93.9	0.1
Georgia	87.2	0.2	85.5	0.2
Hawaii	70.1	0.6	91.5	0.4
Idaho	87.2	0.3	89.7	0.3
Illinois	88.4	0.1	90.1	0.1
Indiana	92.8	0.1	89.7	0.2
Iowa	92.2	0.2	90.8	0.2
Kansas	92.4	0.2	90.4	0.2
Kentucky	79.8	0.3	84.6	0.2
Louisiana	38.3	0.3	86.4	0.4
Maine	81.7	0.3	84.4	0.4
Maryland	91.0	0.2	92.3	0.2
Massachusetts	89.3	0.2	89.5	0.2
Michigan	85.7	0.1	88.3	0.1
Minnesota	88.5	0.1	90.1	0.1
Mississippi	70.6	0.4	75.2	0.4
Missouri	86.7	0.2	89.7	0.2
Montana	79.4	0.4	86.6	0.4
Nebraska	92.1	0.2	91.8	0.2
Nevada	92.2	0.2	96.5	0.2
New Hampshire	83.7	0.4	87.8	0.4
New Jersey	89.4	0.2	94.1	0.1
New Mexico	70.3	0.4	89.4	0.3
New York	88.0	0.1	90.6	0.1
North Carolina	85.5	0.2	84.9	0.2
North Dakota	65.7	0.6	88.3	0.5
Ohio	90.6	0.1	90.8	0.1
Oklahoma	76.3	0.2	88.2	0.2
Oregon	89.7	0.2	92.3	0.2
Pennsylvania	89.0	0.1	90.9	0.1
Rhode Island	91.5	0.5	91.3	0.5
South Carolina	73.5	0.3	84.0	0.3

Table 4 (continued). Coverage and Agreement Rates by State: South Dakota to Wyoming

State	Coverage %	Coverage S.E.	Agreement %	Agreement S.E.
South Dakota	62.5	0.6	89.3	0.5
Tennessee	86.6	0.2	84.1	0.2
Texas	88.2	0.1	92.4	0.1
Utah	88.9	0.3	94.8	0.2
Vermont	83.1	0.5	83.5	0.5
Virginia	88.1	0.2	88.7	0.2
Washington	90.5	0.2	91.8	0.2
West Virginia	20.1	0.4	82.0	0.8
Wisconsin	90.2	0.1	88.7	0.1
Wyoming	80.5	0.7	93.0	0.5

Sources: U.S. Census Bureau: 2019 American Community Survey; Black Knight, Inc.: 2019 and 2021 Assessment and Parcel Boundary Files; DRB Approval Number: CBDRB-FY23-CES019-014

RQ7. When disagreements occur, does the composite file systematically report larger (or smaller) acreage values than the ACS responses?

Finally, Table 5 presents a confusion matrix for the full ACS sample to understand *how* the composite file and the ACS responses tend to differ when they do differ.

Table 5. Confusion Matrix Comparing Reported ACS Values to their Corresponding Administrative Record (AR) Values on the Composite File

	AR Value: Missing	AR Value: 0 to <1 acre	AR Value: 1 to <10 acres	AR Value: 10+ acres
ACS Value: 0 to <1 acres	12.2%	63.3%	2.4%	0.8%
ACS Value: 1 to <10 acres	2.4%	3.5%	10.2%	0.6%
ACS Value: 10+ acres	0.9%	0.2%	0.7%	2.7%

Sources: U.S. Census Bureau: 2019 American Community Survey; Black Knight, Inc.: 2019 and 2021 Assessment and Parcel Boundary Files; DRB Approval Number: CBDRB-FY23-CES019-014

When the composite file has non-missing data, the most likely disagreement is that the ACS response was in the 1 to <10 acres category while the composite file recorded a value of less than one acre. This cell accounted for 3.5% of all cases. This provides indirect evidence of the "rounding up" issue that may cause some ACS respondents on small parcels to report having at least one acre. However, the opposite disagreement (ACS response of less than one acre and composite file recording of 1 to <10 acres) occurred 2.4% of the time, suggesting that disagreements can occur for other reasons. In addition, 0.8% of cases were ones in which the ACS response was under one acre, while the composite file recorded a value of at least ten acres. These disagreements provide indirect evidence that the composite file may report the

acreage of a large parcel that has been subdivided into multiple housing units with small lots. The simulation analysis, discussed in the next section, provides more detail on how acreage estimates might change, across a variety of subgroups, if non-missing information from the composite file replaces ACS responses.

4.2 Simulation Results: Adaptive Design vs Production

4.2.1 Effect on ACS Acreage Estimates Results

RQ8. How might using AR acreage data affect tabular estimates when compared to estimates calculated using the current ACS methodology?

Below is a high level summary of results from Tables 6 through 9 below.

Across a variety of demographic, geographic, and structural characteristics there are certain trends that appeared in the simulation. Including the District of Columbia, nearly half of the states (24) had a statistically significant increase in the proportion of units on under an acre (ACR=1), and for the remaining states, there were no statistically significant differences between the production and simulation results.

The proportion of units on the largest lots (ten acres or more) increased for 31 states, and the remaining states showed no significant difference. This is important to note because one of the main uses for the acreage item is to identify housing units on large lots. The Department of Housing and Urban Development (HUD) and the Bureau of Economic Analysis both subset out these units for their calculations (e.g., HUD's Fair Market Rents). While the trends in aggregate indicate that the simulated data tends to have more housing units reporting their lot is under an acre, this is mostly driven by the single-family detached structure type. This is by far the largest category of housing units in the country, so any trend with those housing units can mask other impacts.

When looking at mobile homes and single-family attached houses, both had a statistically significant decrease in the percentage of housing units on less than an acre, which runs counter to the overall trend. This corresponds with a statistically significant increase in both the *1 to 9.9 acres* and *10 or more acres* categories for both types of structures (mobile homes and single-family attached houses). In the simulated data, 8% of mobile homes were on ten or more acres, as opposed to the 2019 production ACS where 6.2% of mobile homes were on these large lots. Single-family attached houses also had an increase, from 0.6% to 1.4% of this type of structure now being on ten or more acres.

The above breakdown is also apparent based on tenure categories. Rented households (both with cash rent and those occupied without paid rent) and those owned without a mortgage had increases in the number of units on ten or more acres, but both categories of rented units had much larger changes in their proportions. Units with cash rent went from 1.8% of their units on

ten or more acres to 3.4%. Units not paying cash rent went from 10% on large lots to 13.1%. For owned units, the trend mimicked the overall trend, with a shift from units on 1 to 9.9 acres to units on under an acre.

Many housing items have differing item nonresponse rates based on the tenure (owned vs. rented of the unit), and acreage does as well. Owned units had an item nonresponse rate of 3% in the 2019 ACS for acreage, while rented units had a nonresponse rate of 4.8%. Using administrative data can help mitigate any bias that the increased imputation rates for rented units introduce.

Table 6. Comparison of Acreage Values by Respondent Age and Sex: Simulation (SIM) vs. Production (PROD); (ACR = 1: Less than 1 Acre, ACR = 2: 1 to 9.9 Acres, ACR = 3: 10 or More Acres)

Respondent Demographic Characteristics	ACR = 1 SIM % (S.E)	ACR = 1 PROD % (S.E)	ACR = 2 SIM % (S.E)	ACR = 2 PROD % (S.E)	ACR = 3 SIM % (S.E)	ACR = 3 PROD % (S.E)
AGE	-	-	-	-	-	-
Under 30 years old	84.9 (0.1)	84.7 (0.2)	11.6* (0.1)	12.8 (0.2)	3.5* (0.1)	2.5 (0.1)
30 to 39 years old	84.7* (0.1)	83.7 (0.1)	12.2* (0.1)	13.8 (0.1)	3.0* (<0.1)	2.5 (<0.1)
40 to 49 years old	82.4* (0.1)	81.0 (0.1)	14.2* (0.1)	15.9 (0.1)	3.3* (<0.1)	3.0 (<0.1)
50 to 59 years old	78.8* (0.1)	77.3 (0.1)	16.8* 0.1)	18.5 (0.1)	4.4* (<0.1)	4.2 (<0.1)
60 to 69 years old	76.9* (0.1)	75.7 (0.1)	17.7* (0.1)	19.0 (0.1)	5.4* (<0.1)	5.3 (<0.1)
70 years old or older	77.1* (0.1)	76.3 (0.1)	16.6* (0.1)	17.9 (0.1)	6.3* (0.1)	5.8 (<0.1)
SEX	-	-	-	-	-	-
Female	81.4* (0.1)	80.3 (0.1)	14.5* (<0.1)	16.0 (0.1)	4.0* (<0.1)	3.7 (<0.1)
Male	78.7* (0.1)	77.6 (0.1)	16.3* (0.1)	17.7 (0.1)	5.0* (<0.1)	4.6 (<0.1)

Sources: U.S. Census Bureau: 2019 American Community Survey; Black Knight, Inc.: 2019 and 2021 Assessment and Parcel Boundary Files; DRB Approval Number: CBDRB-FY23-CES019-014

Note: Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two-tailed t-test at the α=0.1 level. This table includes estimates from all states, the District of Columbia, and Puerto Rico.

Table 7. Comparison of Acreage Values by Respondent Hispanic Origin and Race: Simulation (SIM) vs. Production (PROD) National and Puerto Rico; (ACR = 1: Less than 1 Acre, ACR = 2: 1 to 9.9 Acres, ACR = 3: 10 or More Acres)

Respondent Demographic Characteristics	ACR = 1 SIM % (S.E)	ACR = 1 PROD % (S.E)	ACR = 2 SIM % (S.E)	ACR = 2 PROD % (S.E)	ACR = 3 SIM % (S.E)	ACR = 3 PROD % (S.E)
HISPANIC ORIGIN	-	-	-	-	-	-
Hispanic or Latino	92.1* (0.1)	91.8 (0.1)	6.5* (0.1)	7.3 (0.1)	1.4* (<0.1)	0.9 (<0.1)
Not Hispanic or Latino	78.3* (<0.1)	77.1 (0.1)	16.8* (<0.1)	18.3 (<0.1)	5.0* (<0.1)	4.6 (<0.1)
RACE	-	-	-	-	-	-
White alone	77.4* (<0.1)	76.5 (0.1)	17.4* (<0.1)	18.6 (<0.1)	5.2* (<0.1)	4.9 (<0.1)
Black or African American alone	90.9* (0.1)	87.8 (0.1)	7.8* (0.1)	11.3 (0.1)	1.3* (<0.1)	0.9 (<0.1)
American Indian and Alaska Native alone	74.2 (0.5)	75.2 (0.4)	18.3* (0.5)	20.3 (0.4)	7.5* (0.3)	4.5 (0.2)
Asian alone	93.7* (0.1)	93.3 (0.1)	5.3* (0.1)	6.1 (0.1)	1.0* (<0.1)	0.6 (<0.1)
Native Hawaiian and Other Pacific Islander alone	89.7 (1.0)	89.7 (1.0)	7.6 (0.9)	8.0 (0.9)	2.7 (0.5)	2.3 (0.5)
Some Other Race alone	93.0* (0.2)	92.5 (0.2)	5.6* (0.1)	6.8 (0.2)	1.4* (0.1)	0.7 (0.1)
Two or More Races	85.4* (0.3)	84.0 (0.2)	11.6* (0.2)	13.3 (0.2)	3.0 (0.1)	2.7 (0.1)

Sources: U.S. Census Bureau: 2019 American Community Survey; Black Knight, Inc.: 2019 and 2021 Assessment and Parcel Boundary Files; DRB Approval Number: CBDRB-FY23-CES019-014

Note: Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two-tailed t-test at the α=0.1 level. For Race: All race categories represent single-race responses, except the "Two or More Races" category. This table includes estimates from all states, the District of Columbia, and Puerto Rico.

Table 8. Comparison of Housing Characteristics by Acreage Values: Simulation (SIM) vs. Production (PROD); (ACR=1:less than 1 Acre, ACR=2: 1 to 9.9 Acres, ACR=3:10 or more Acres)

Responding Household Characteristics	SIM ACR=1 % (S.E.)	PROD ACR=1 % (S.E.)	SIM ACR=2 % (S.E.)	PROD ACR=2 % (S.E.)	SIM ACR=3 % (S.E.)	PROD ACR=3 % (S.E.)
OCCUPANCY STATUS	-	-	-	-	-	-
Occupied	80.0* (<0.1)	78.9 (<0.1)	15.5* (<0.1)	16.9 (<0.1)	4.5* (<0.1)	4.1 (<0.1)
Vacant	74.6* (0.2)	77.7 (0.1)	17.2* (0.1)	15.7 (0.1)	8.3* (0.1)	6.5 (0.1)
BUILDING TYPE	-	-	-	-	-	-
Mobile home	66.1* (0.1)	69.0 (0.1)	25.9* (0.1)	24.8 (0.1)	8.0* (0.1)	6.2 (0.1)
Single-family, detached	79.3* (<0.1)	78.2 (<0.1)	15.7* (<0.1)	17.2 (<0.1)	5.0* (<0.1)	4.6 (<0.1)
Single-family, attached	93.5* (0.1)	95.3 (0.1)	5.1* (0.1)	3.9 (0.1)	1.4* (<0.1)	0.8 (<0.1)
YEAR BUILT	-	-	-	-	-	-
2018 and 2019	81.6 (0.4)	82.6 (0.4)	13.2 (0.4)	13.3 (0.4)	5.2* (0.2)	4.2 (0.2)
2010 to 2017	79.0 (0.2)	79.2 (0.2)	15.7* (0.2)	16.1 (0.2)	5.2* (0.1)	4.7 (0.1)
2000 to 2009	76.6 (0.1)	76.6 (0.1)	18.3* (0.1)	18.6 (0.1)	5.2* (0.1)	4.8 (<0.1)
1970 to 1999	75.5* (0.1)	75.2 (0.1)	19.3* (0.1)	20.2 (0.1)	5.2* (<0.1)	4.6 (<0.1)
1940 to 1969	86.6* (0.1)	85.1 (0.1)	10.2* (0.1)	12.2 (0.1)	3.2* (<0.1)	2.7 (<0.1)
1939 or earlier	80.2* (0.1)	79.2 (0.1)	11.9* (0.1)	13.6 (0.1)	7.8* (0.1)	7.2 (0.1)
TENURE	-	-	-	-	-	-
Owned with mortgage or loan	81.5* (0.1)	79.5 (0.1)	15.5* (<0.1)	17.4 (0.1)	3.0* (<0.1)	3.2 (0.1)
Owned free and clear	74.5* (0.1)	73.3 (0.1)	18.5* (0.1)	20.2 (0.1)	6.9* (<0.1)	6.6 (<0.1)
Rent with payment	87.6* (0.1)	89.3 (0.1)	8.9 (0.1)	8.8 (0.1)	3.4* (0.1)	1.8 (<0.1)
Occupied without payment	67.7* (0.4)	69.6 (0.3)	19.2* (0.3)	20.4 (0.3)	13.1* (0.3)	10.0 (0.2)

Sources: U.S. Census Bureau: 2019 American Community Survey; Black Knight, Inc.: 2019 and 2021 Assessment and Parcel Boundary Files; DRB Approval Number: CBDRB-FY23-CES019-014

Note: Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two-tailed t-test at the at the α=0.1 level. This table includes estimates from all states, the District of Columbia, and Puerto Rico.

Table 9. Percentage of Sampled Housing Units in each Acreage Category by State: Alabama to Massachusetts: Simulation (SIM) vs. Production (PROD)
(ACR=1: Less than 1 Acre, ACR=2: 1 to 9.9 Acres, ACR=3: 10 or More Acres)

State	SIM ACR=1 % (S.E.)	PROD ACR=1 % (S.E.)	SIM ACR=2 % (S.E.)	PROD ACR=2 % (S.E.)	SIM ACR=3 % (S.E.)	PROD ACR=3 % (S.E.)
Alabama	68.2*	66.3	24.9*	27.5	6.9*	6.2
	(0.3)	(0.4)	(0.3)	(0.3)	(0.2)	(0.2)
Alaska	67.7	66.4	29.0	30.1	3.3	3.5
	(0.7)	(0.8)	(0.8)	(0.8)	(0.4)	(0.3)
Arizona	88.8*	88.4	9.9*	10.8	1.3*	0.9
	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(<0.1)
Arkansas	63.9*	65.4	26.8*	25.6	9.3	9.0
	(0.4)	(0.4)	(0.5)	(0.4)	(0.3)	(0.3)
California	89.9*	91.5	7.7*	7.0	2.4*	1.5
	(0.1)	(0.1)	(0.1)	(0.1)	(<0.1)	(<0.1)
Colorado	84.5*	83.9	10.4*	11.4	5.1*	4.7
	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)
Connecticut	70.3*	68.3	27.1*	29.8	2.6*	2.0
	(0.4)	(0.4)	(0.4)	(0.4)	(0.2)	(0.1)
Delaware	87.1	86.3	10.3*	12.2	2.7*	1.5
	(0.5)	(0.5)	(0.5)	(0.5)	(0.3)	(0.2)
District of Columbia	99.6*	97.8	0.4*	2.0	0.0	0.2
	(0.2)	(0.3)	(0.2)	(0.3)	(0)	(0.1)
Florida	89.1	89.2	9.3*	9.7	1.6*	1.1
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Georgia	72.4*	70.8	22.8*	25.3	4.7*	4.0
	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)
Hawaii	85.2*	86.8	12.4	12.0	2.4*	1.3
	(0.6)	(0.5)	(0.5)	(0.5)	(0.1)	(0.2)
Idaho	74.1	74.2	19.4	20.4	6.5*	5.5
	(0.5)	(0.5)	(0.5)	(0.5)	(0.2)	(0.3)
Illinois	87.0	86.8	9.2*	10.6	3.8*	2.6
	(0.1)	(0.1)	(0.1)	(0.1)	(0.3)	(0.1)
Indiana	77.7*	76.1	16.6*	18.7	5.7*	5.2
	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)
Iowa	77.9	77.7	13.7*	15.1	8.4*	7.3
	(0.3)	(0.2)	(0.2)	(0.2)	(0.1)	(0.2)
Kansas	80.7*	80.0	11.3*	12.9	8.0*	7.1
	(0.3)	(0.3)	(0.2)	(0.3)	(0.2)	(0.2)
Kentucky	66.5*	65.4	22.2*	24.4	11.3*	10.1
	(0.4)	(0.3)	(0.3)	(0.4)	(0.2)	(0.2)
Louisiana	76.2	76.1	19.7	20.0	4.1	3.9
	(0.3)	(0.3)	(0.3)	(0.3)	(0.2)	(0.2)
Maine	47.4	47.0	41.2	42.3	11.3	10.7
	(0.5)	(0.6)	(0.6)	(0.6)	(0.3)	(0.3)
Maryland	84.0*	81.5	13.7*	16.4	2.3*	2.0
	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)
Massachusetts	79.8*	78.3	18.1*	20.3	2.1*	1.5
	(0.3)	(0.2)	(0.3)	(0.2)	(0.1)	(0.1)

Table 9 (continued). Percentage of Sampled Housing Units in each Acreage Category by State: Michigan to Texas: Simulation (SIM) vs. Production (PROD)
(ACR=1: Less than 1 Acre, ACR=2: 1 to 9.9 Acres, ACR=3: 10 or More Acres)

State	SIM ACR=1 % (S.E.)	PROD ACR=1 % (S.E.)	SIM ACR=2 % (S.E.)	PROD ACR=2 % (S.E.)	SIM ACR=3 % (S.E.)	PROD ACR=3 % (S.E.)
Michigan	74.8	74.9	17.8	18.2	7.4*	6.9
	(0.2)	(0.2)	(0.1)	(0.2)	(0.1)	(0.1)
Minnesota	73.5*	72.4	16.7*	18.0	9.8	9.6
	(0.3)	(0.3)	(0.2)	(0.2)	(0.1)	(0.1)
Mississippi	57.5*	59.9	32.1	31.7	10.4*	8.4
	(0.4)	(0.4)	(0.4)	(0.5)	(0.3)	(0.2)
Missouri	74.3	74.3	15.7*	16.2	10.0*	9.5
	(0.2)	(0.3)	(0.2)	(0.2)	(0.2)	(0.2)
Montana	63.1	64.0	22.1	22.2	14.8	13.8
	(0.7)	(0.6)	(0.6)	(0.6)	(0.5)	(0.4)
Nebraska	82.0	82.1	9.0*	10.6	9.0*	7.4
	(0.3)	(0.3)	(0.3)	(0.2)	(0.2)	(0.2)
Nevada	92.2	91.8	6.4*	7.0	1.4	1.2
	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)
New Hampshire	52.0	51.4	39.6	40.9	8.4	7.7
	(0.6)	(0.6)	(0.7)	(0.6)	(0.4)	(0.3)
New Jersey	89.8*	88.1	9.3*	11.1	0.9	0.9
	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)
New Mexico	79.2	78.9	17.1	17.7	3.7	3.4
	(0.4)	(0.4)	(0.4)	(0.4)	(0.2)	(0.2)
New York	78.8*	76.4	16.1*	18.7	5.1*	4.9
	(0.2)	(0.2)	(0.1)	(0.2)	(0.1)	(0.1)
North Carolina	73.4*	71.1	22.1*	25.3	4.5*	3.5
	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)
North Dakota	74.0	73.8	12.6*	14.3	13.3*	11.9
	(0.5)	(0.6)	(0.5)	(0.6)	(0.4)	(0.4)
Ohio	79.3*	77.8	16.3*	18.2	4.3*	4.0
	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)
Oklahoma	72.1	72.3	18.5	18.7	9.4*	9.0
	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)
Oregon	79.3*	80.2	14.7	14.4	6.0*	5.4
	(0.3)	(0.3)	(0.3)	(0.3)	(0.2)	(0.2)
Pennsylvania	78.9*	76.7	16.1*	18.5	5.0	4.9
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Rhode Island	82.2*	79.6	15.4*	18.8	2.4*	1.6
	(0.8)	(0.7)	(0.7)	(0.6)	(0.3)	(0.3)
South Carolina	73.9*	73.2	21.9*	23.5	4.2*	3.3
	(0.3)	(0.3)	(0.3)	(0.3)	(0.1)	(0.1)
South Dakota	73.2	74.2	14.8	14.7	12.0	11.1
	(0.6)	(0.6)	(0.6)	(0.5)	(0.4)	(0.4)
Tennessee	68.0*	66.2	24.2*	27.1	7.7*	6.8
	(0.3)	(0.3)	(0.3)	(0.3)	(0.2)	(0.2)
Texas	84.6*	84.2	11.7*	12.1	3.7	3.7
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)

Table 9 (continued). Percentage of Sampled Housing Units in each Acreage Category by State: Utah to Wyoming: Simulation (SIM) vs. Production (PROD)
(ACR=1: Less than 1 Acre, ACR=2: 1 to 9.9 Acres, ACR=3: 10 or More Acres)

State	SIM ACR=1 % (S.E.)	PROD ACR=1 % (S.E.)	SIM ACR=2 % (S.E.)	PROD ACR=2 % (S.E.)	SIM ACR=3 % (S.E.)	PROD ACR=3 % (S.E.)
Utah	90.5*	89.1	8.2*	9.7	1.3	1.1
	(0.3)	(0.3)	(0.3)	(0.3)	(0.1)	(0.1)
Vermont	43.5	43.5	33.9	34.3	22.7	22.2
	(0.8)	(0.7)	(0.7)	(0.7)	(0.6)	(0.6)
Virginia	74.2*	72.9	20.2*	21.9	5.6*	5.2
	(0.2)	(0.2)	(0.2)	(0.2)	(0.1)	(0.1)
Washington	79.3*	80.3	17.1*	16.4	3.5	3.4
	(0.2)	(0.2)	(0.3)	(0.2)	(0.1)	(0.1)
West Virginia	64.9	64.7	26.1	26.0	8.9	9.3
	(0.5)	(0.5)	(0.4)	(0.4)	(0.3)	(0.4)
Wisconsin	68.1*	67.0	21.8*	23.1	10.0	9.9
	(0.2)	(0.3)	(0.2)	(0.3)	(0.1)	(0.1)
Wyoming	69.4	69.1	19.6	20.1	11.0	10.8
	(0.9)	(0.9)	(0.9)	(1.0)	(0.5)	(0.6)

Sources: U.S. Census Bureau: 2019 American Community Survey; Black Knight, Inc.: 2019 and 2021 Assessment and Parcel Boundary Files; DRB Approval Number: CBDRB-FY23-CES019-014
Note: Standard errors are in parentheses. An asterisk (*) indicates a statistically significant result based on a two-tailed t-test at the α=0.1 level.

4.2.2 Effect on ACS Agricultural Sales Estimates Results

RQ9. How does using AR acreage data affect tabular estimates for Agricultural Sales when compared to estimates calculated using the current ACS methodology?

Because the agricultural sales (AGS) question is asked based on the lot size categories, we evaluated the impact of the simulation on agricultural sales estimates. The effect on the AGS question was quite substantial during this simulation but should be mitigated in production.

In this test, all cases that were moved into the AGS universe in the simulation, moving from under an acre (ACR=1) to an acre or more (ACR=2 or ACR=3), needed to have their AGS values imputed. This drove up the imputation rate for agricultural sales, as expected. We expect the imputation rate to come down as this work is moved into production since respondents will be asked the AGS question if either their responded acreage value or the administrative data value indicates the address is on an acre or more. This was one of the major reasons that the adaptive approach was adopted, to ensure that the AGS question would still be asked, regardless of the method that acreage data is gathered.

In addition, the number of units reporting agricultural sales increased dramatically. The reason for this increase is that the hot deck matrices for the AGS imputation uses the size of the lot (ACR=2 vs ACR=3) as one of the correlates. Units on ten or more acres (ACR=3) are more likely to have agricultural sales than those on 1 to 9.9 acres, as 31 states saw a statistically significant increase in their proportions of units with greater than ten acres. Thus, the imputed cases were

drawing from a donor pool that was more likely to have agricultural sales. This is most likely an artifact of the simulation and will not be replicated in production, since in production the AGS question will be asked of respondents and not imputed.

It is possible that the number of units reporting agricultural sales will increase as this work is adopted, as many states saw an increase eligibility for being asked the AGS questions (i.e., units on ten or more acres). We have confidence in the accuracy of the acreage administrative data, so if there also is an increase in agricultural sales, it should also reflect accurate data. We, of course, will not know the effects until production implementation.

5. CONCLUSION

As household surveys face declining response rates, researchers at the Census Bureau have increasingly analyzed ways to supplement household survey data with sources of AR to provide quality information about the U.S. population and economy. The Census Bureau is actively looking for ways to use AR data to enhance the ACS and reduce the burden placed on our respondents, improve data quality, and create blended data products to meet data user needs. Prior research had identified the acreage item as a likely candidate for replacement or extensive supplementation with AR sources. This report described the creation and testing of an adaptive design to produce acreage information from a combination of survey responses and AR data from property tax assessments provided by Black Knight, Incorporated.

One key complication with which our research grappled was the conceptual misalignment between the *housing units* sampled by the ACS and the *land parcels* about which information is recorded by county assessor's offices. In some cases, this resulted in the inability to link acreage information to housing-unit addresses, resulting in imperfect coverage; in other cases (e.g., a large parcel that is subdivided into several mobile homes or townhouses), this resulted in substantial disagreement between survey-based and AR-based acreage information. We also noted a small subset of counties that appeared to provide data of especially low reliability.

These complications prohibited a full replacement of survey responses with AR information. Instead, we developed an adaptive design, in which the acreage question would be removed for the majority of (but not all) respondents and tested its performance in two ways. First, we built a composite file in which we optimally combined the various sources of AR data provided by Black Knight—across vintages and data types (address-based, geospatial) with suppressions for poor-quality counties—and analyzed its coverage of, and agreement with, ACS acreage responses from the 2019 survey. Second, we simulated how a variety of 2019 acreage estimates would change if we replaced the 2019 acreage value from the final ACS production file with the value from the composite file, where the latter was non-missing and where the response mode was internet or CAPI.

Our first analysis yielded high, but imperfect, rates of coverage and agreement. We matched approximately 84.5% of 2019 ACS housing units to a valid AR, and 90.2% of these AR values matched the corresponding ACS survey response. Coverage and agreement rates varied relatively little across demographic subgroups, although housing units headed by an individual identifying as American Indian or Alaska Native had a substantially lower coverage rate and slightly lower agreement rate than the national average. Coverage rates varied substantially across building types—from 42.8% for mobile homes to 62.7% for single-family attached houses to 90.6% for single-family detached houses—but agreement rates varied less by building type. Rented houses and houses outside of metro areas also had a substantially lower coverage rate, and a slightly lower agreement rate, than owned houses or houses inside metro areas. Outside of several small-state outliers that had notably low coverage rates, coverage and agreement rates varied little across states.

The simulations confirm several known issues with respondent-reported values, mainly having to do with their propensity to round estimates. Prior studies on the acreage data collected from the American Housing Survey and ACS indicated that respondents most likely were over-reporting the size of their lots, most commonly rounding up to a single acre. This simulated acreage data includes a statistically significant increase in the number of occupied units on less than an acre (80% compared to 78.9%) which is also the most prevalent type of unit throughout the country. This increase comes alongside a decrease in the units reporting 1 to 9.9 acres.

One result that will be monitored moving forward is the statistically significant increase in the estimates for the proportion of housing units residing on ten or more acres. This was across all structure types, but mobile homes and single-family attached houses saw the largest proportional increases. All states either showed no statistical difference between the production and simulation data or had a statistically significant increase in the number of these units. This could be a function of the conceptual error, mentioned above, or confusion on the respondent's part when determining their true lot size.

While the research points to some possible impacts on ACS estimates, we believe that the benefits of using administrative data in place of survey responses for acreage outweigh any of the possible downfalls. Incorporating this data will help reduce respondent burden and potentially increase data quality for a vast majority of housing units. This change will also bring the ACS in line with the other Census survey that also collects this information, the AHS, which has been using a similar methodology for nearly a decade. We recommend implementing administrative records data for acreage into full ACS production according to the methodology outlined in this report. We will continuously monitor the use of administrative acreage data and we may update the methodology when necessary or possible.

6. REFERENCES

Binder, Ariel J., Emily Molfino, and John Voorheis. 2022. "Comparing the 2019 American Housing Survey to Contemporary Sources of Property Tax Records: Implications for Survey Efficiency and Quality." *Proceeding of the 2022 Joint Statistical Meetings*, *Survey Research Methods Section*: 1688-1717.

Brummet, Quentin. 2015. "Matching Addresses between Household Surveys and Commercial Data." CARRA Working Paper Series WP-2015-04. https://census.gov/content/dam/Census/library/working-papers/2015/adrm/carra-wp-2015-04.pdf

Clark, Sandra, R. Chase Sawyer, Amanda Klimek, Christopher Mazur, William Chapin, and Ellen Wilson. 2018. Housing Administrative Record Simulation. Washington, DC: U.S. Census Bureau. https://www.census.gov/library/working-papers/2018/acs/2018_Clark_01.html

Dillon, Michaela. 2019. Use of Administrative Records to Replace or Enhance Questions about Housing Characteristics on the American Community Survey. Washington, DC: U.S. Census Bureau. https://www.census.gov/library/working-papers/2019/acs/2019_Dillon_01.html

Molfino, Emily. 2021a. "Imputing Lot Size with Property Tax Data." US Department of Housing and Urban Development Office of Policy Development and Research. https://www.census.gov/content/dam/Census/programs-surveys/ahs/working-papers/Imputing-Lot-Size-Property-Tax-Data.pdf

Molfino, Emily. 2021b. "Imputing Year Built with Property Tax Data." US Department of Housing and Urban Development Office of Policy Development and Research. https://www.census.gov/content/dam/Census/programs-surveys/ahs/working-papers/Imputing-Year-Built-Property-Tax-Data.pdf

Molfino, Emily, Gizem Korkmaz, Sallie A. Keller, Aaron Schroeder, Stephanie Shipp, and Daniel H. Weinberg. 2017. "Can Administrative Housing Data Replace Survey Data?" Cityscape 19.1: 265-292. https://www.huduser.gov/portal/periodicals/cityscpe/vol19num1/ch16.pdf

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

U.S. Census Bureau. 2020. "Agility in Action 3.0: A Snapshot of Enhancements to the ACS." https://www.census.gov/programs-surveys/acs/methodology/agility-in-action/agility-in-action-3.html

Weinberg, Daniel H. 2015. "Data Sources for US Housing Research, Part 2: Private Sources, Administrative Records, and Future Directions." Cityscape 17.1: 191-206. https://www.huduser.gov/portal/periodicals/cityscpe/vol17num1/ch14.pdf