



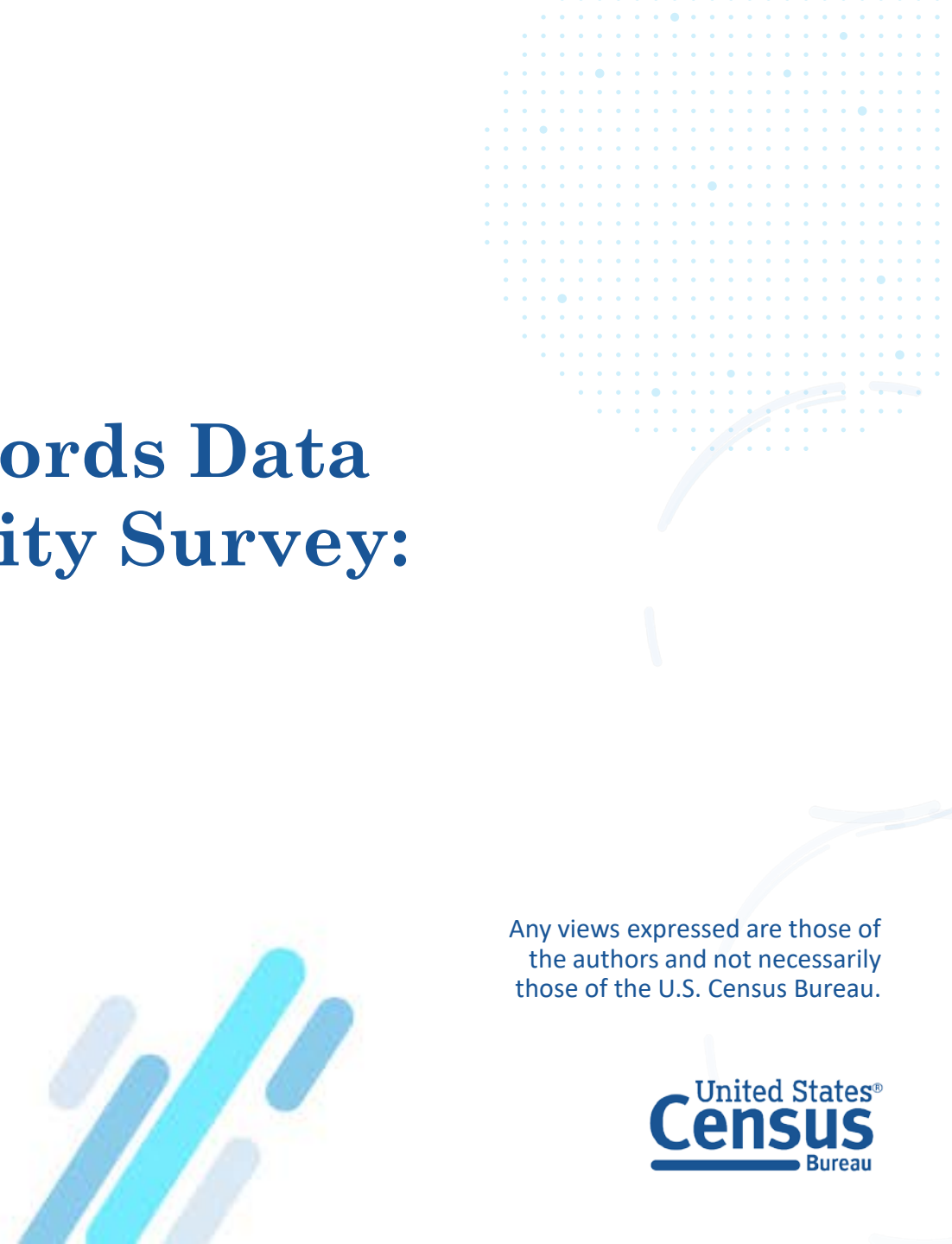
Using Administrative Records Data in the American Community Survey: Overview and Updates

2025 ACS Data Users Conference

Dorothy Barth

U.S. Census Bureau

May 29, 2025



Any views expressed are those of
the authors and not necessarily
those of the U.S. Census Bureau.

Setting the Stage

- Discussion on Administrative Records Data
- Review of the Gates to Implementation for the ACS
- Updates on Current Production and Research Projects

Administrative Records Data

- Administrative records (Ad Rec) data are data collected from government agencies to administer programs or provide services.
- These data are distinct from the data collected exclusively for statistical purposes, like those the Census Bureau produces under the authority of Title 13 U.S. Code.
- Sometimes, commercial vendor data or free, publicly available data are also broadly referred to as Ad Rec data.

Administrative Records Data

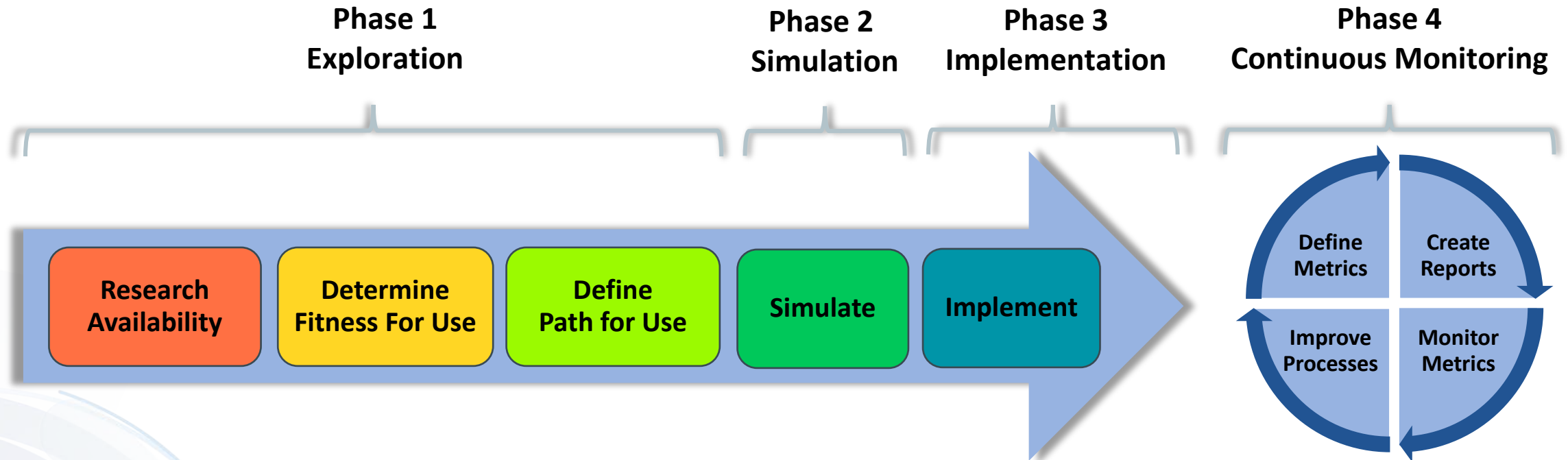
Examples of Ad Rec data being used by the ACS:

- Social Security Data
- IRS Tax Form Data
- Data from the USPS indicating when and why mailings are undeliverable as addressed (UAA)
- Commercial Data (ex.: phone numbers and property data)
- USDA Cropland Data (publicly available)

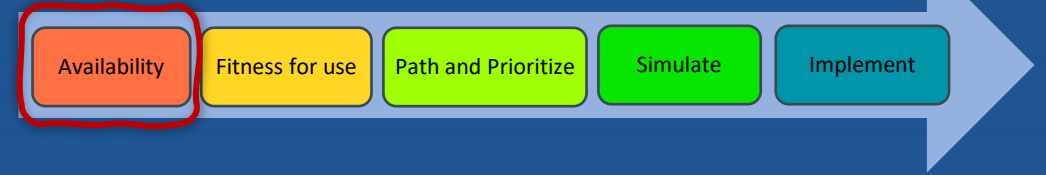
Why use administrative records data?

- Obligated by law (Title 13, U.S. Code) to use existing information in lieu of direct inquiry.
- Allowed to use these data for statistical purposes.
- Committed to reducing respondent burden and improving data quality.

Implementation Phases and Gates



Gates to Implementation



ACS DATA

- Mandated Use
- ACS Product
- Geographic Need

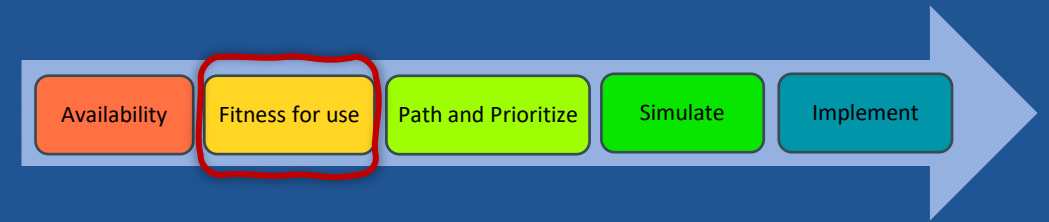
AD REC DATA

- Data Source
- Years Available
- Geography Available

CONTRACT INFO.

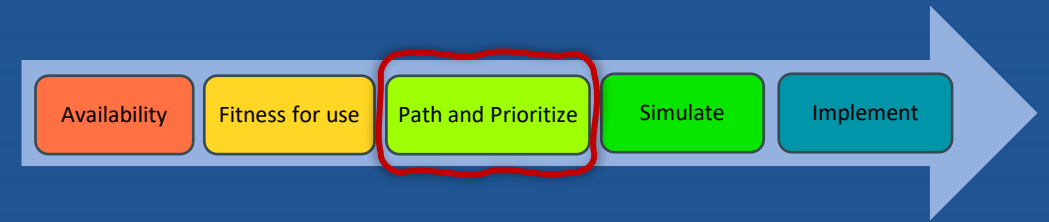
- Annual Cost
- Delivery Frequency
- Link to the ACS sample

Gates to Implementation



COVERAGE	How much Ad Rec data exists at the lowest geographic levels?
CONCEPTUAL ALIGNMENT	How well do the Ad Rec data match when linked to the ACS sample?
TEMPORAL ALIGNMENT	How well does the Ad Rec data correspond to the period referenced in the ACS question?
DISCLOSURE AVOIDANCE	Are individuals or households easily identifiable?
RISK OVER TIME	Are there risks with the contract agreement or timing of data delivery?

Gates to Implementation

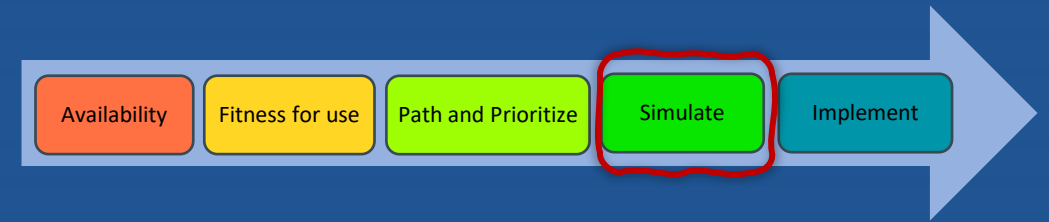


Survey Life Cycle



OPTIMIZING OPERATIONS	Use administrative data to inform ACS operations and processes.
REPLACING QUESTIONS	Remove the question from the survey.
PARTIALLY REPLACING QUESTIONS	Ask the survey question of some, but not all, respondents.
EDITING AND IMPUTING DATA	Keep the question on the survey. Use Ad Rec data to replace or edit survey responses and impute missing responses.
ENHANCING DATA PRODUCTS	Use Ad Rec data to create new products or enhance existing data products.

Gates to Implementation



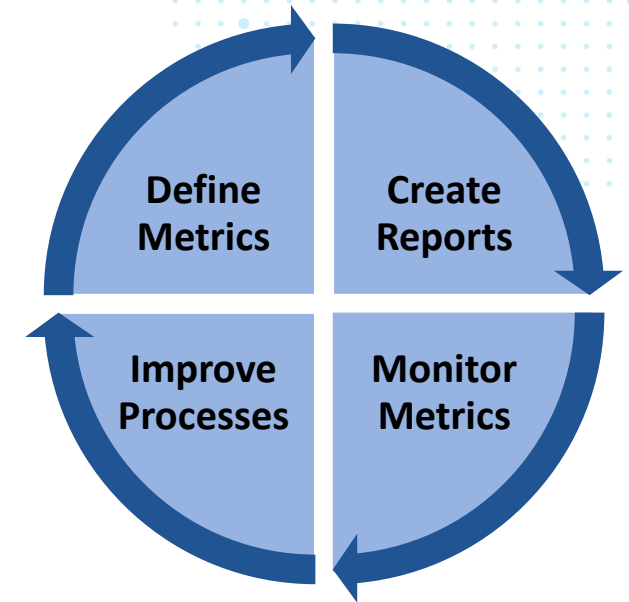
ITEM NONRESPONSE	How much did we reduce or increase item nonresponse?
BIAS	How much did we reduce or increase bias?
RELIABILITY	How much did we reduce or increase the reliability of the ACS data?
DOWNSTREAM IMPACT	Does adding Ad Rec data have a positive or negative impact on other survey items?
OVERALL DATA QUALITY	What is the overall impact to data quality in our published tables and other data products?

Phase 4: Continuous Monitoring

Once a project is implemented in production, we will continuously monitor things like:

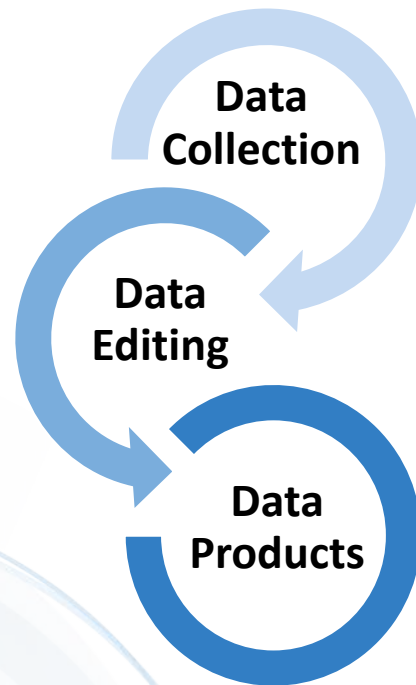
- Business Rules for Ad Rec data use
- Availability and Quality of the Ad Rec data
- Model performance over time
- Outcomes of published survey data

When necessary, we will make updates to our processes to address any concerns that may arise.



Status of Current Administrative Data Use and Research

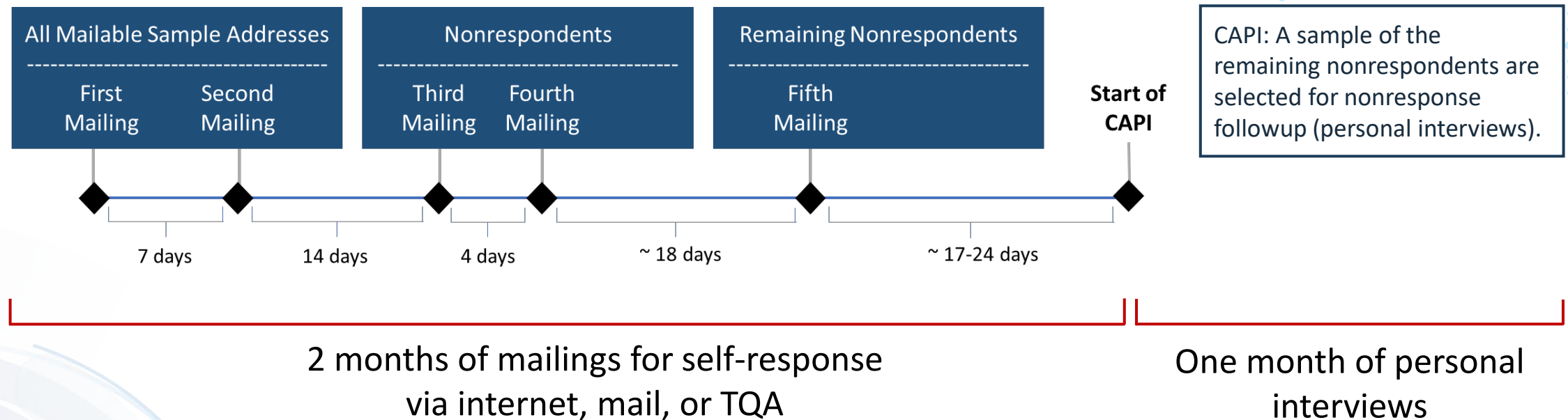
Survey Life Cycle



OPTIMIZING OPERATIONS	CAPI Workload Optimization (2023) M
REPLACING QUESTIONS	Agricultural Sales S
PARTIALLY REPLACING QUESTIONS	Acreage/Lot Size (2024) I Year Built and other Housing Items E
EDITING AND IMPUTING DATA	Sex and Age S
ENHANCING DATA PRODUCTS	Farm Indicator S

E=Exploration **S**=Simulation **I**=Implementation **M**=Monitoring

The ACS Data Collection Strategy



- At the beginning of the CAPI month, we run a model to identify CAPI sample housing units that are likely vacant.
- Cases identified in the top 1% become eligible for “stop work.”
- After two contact attempts, each case identified is removed from the workload.
- The intention is to optimize workload efficiency to allow interviewers more time to collect data from households that are most likely occupied.
- On average, this process stops about 0.15% of the CAPI workload each month.

- About 15 days into the CAPI month, we run the three models to identify more cases to remove from the CAPI workload. The models are run on the remaining incomplete CAPI cases that are NOT likely vacant.
- Each month, after 10 full weekdays and two full weekends, we remove cases from the workload identified by the models.
- On average, this process stops about 3.6% of the CAPI workload each month.

Three models work together to identify cases:

Quality: Responding and non-responding characteristics

- ▶ from administrative records data

Response Propensity: Case-level response propensity models

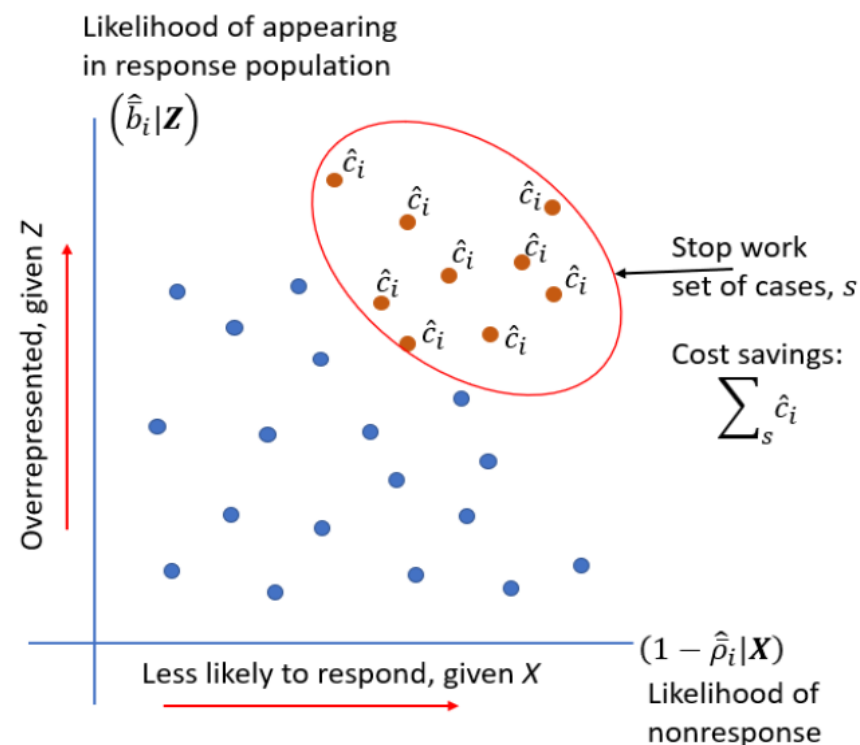
- ▶ from accumulating paradata

Cost: Case-level data collection cost models

- ▶ from accumulating paradata and cost data

How to Identify Which Cases to Stop?

(lower likelihood of responding, well-represented characteristics)



Acreage Question – Began January 2024

PARTIALLY REPLACING
QUESTIONS

The ACS asks the acreage question to people living in houses and mobile homes:

How many acres is this house or mobile home on?

- ☐ Less than 1 acre → *SKIP to question 6a*
- ☐ 1 to 9.9 acres
- ☐ 10 or more acres

Acreage Question – Began January 2024

PARTIALLY REPLACING
QUESTIONS

Beginning with the 2024 survey year, the ACS used third-party lot size data to replace survey answers for the acreage question.

- The acreage question is skipped on the Internet and CAPI data collection instruments when there is high-quality Ad Rec data available.
- About 71% of all responding households were not asked this question in 2024.
- This project is still in the implementation phase because we have not yet completed all production phases and published the data.

Agricultural Sales Question

REPLACING
QUESTIONS

ENHANCING DATA
PRODUCTS

The ACS asks the agricultural sales question to people living in houses and mobile homes on more than one acre:

IN THE PAST 12 MONTHS, what were the actual sales of all agricultural products from this property?

- ☐ None
- ☐ \$1 to \$999
- ☐ \$1,000 to \$2,499
- ☐ \$2,500 to \$4,999
- ☐ \$5,000 to \$9,999
- ☐ \$10,000 or more

Agricultural Sales Question

REPLACING
QUESTIONS

ENHANCING DATA
PRODUCTS

- The only known federal use for the Ag Sales data is by the U.S. Bureau of Economic Analysis (BEA).
- BEA uses the data to create a distinction between farm and non-farm properties.
- The farm designation is used to aid in calculations for their data releases on National Income and Product Accounts.

Agricultural Sales Question

REPLACING
QUESTIONS

ENHANCING DATA
PRODUCTS

- We are working closely with BEA to develop a model that predicts a farm/non-farm designation using Ad Rec data as model inputs.
- We are proposing to publish a new Farm Indicator data product.
- The current proposal is to keep the Ag Sales question on the survey for a while, but to remove it eventually.
- This project is still in the simulation phase.

Sex and Age Questions

EDITING AND
IMPUTING DATA

For each household member, we ask their sex and age.

What is Person 1's sex? *Mark (X) ONE box.*

☐

Male

☐

Female

What is Person 1's age and what is Person 1's date of birth? *For babies less than 1 year old, do not write the age in months. Write 0 as the age.*

Print numbers in boxes.

Age (in years)

Month

Day

Year of birth

Sex and Age Questions

EDITING AND
IMPUTING DATA

- Ad Rec data were used to impute Sex and Age for the 2020 Census.
- The ACS began research on sex and age in 2019. We are updating this research using newer data and applying lessons learned about Ad Rec data since then.
- We are in the simulation phase, but the proposal is to edit responses or impute nonresponses when Ad Rec data is available.

Housing Topic Questions

PARTIALLY REPLACING
QUESTIONS

We are currently in the Exploration Phase for the following topics:

- Year Built
- Number of Rooms/Bedrooms
- Heating/Fuel
- Units in Structure
- Building Type

If the Ad Rec data is sufficient, the data will most likely be used to skip the housing questions on the automated Internet and CAPI instruments, like we are doing for Acreage.

References

Agility in Action 4.0: Responding and Adapting to Our Evolving Nation. (2025). U.S. Census Bureau. <https://www.census.gov/content/dam/Census/programs-surveys/acs/methodology/agility-in-action/Agility-in-Action-4.0-Revision-1.pdf>

Clark, S., Barth, D., Binder, A., Diagne, A., Pharris-Ciurej, N., Voorheis, J., Brassell, E., & Mazur, C. (2023). *Acreage Administrative Records Data Research for the American Community Survey.* U.S. Census Bureau. https://www.census.gov/library/working-papers/2023/acs/2023_Clark_01.html

Keller, A. (2021). *Administrative Record Use in the 2020 Census For Modeling and Processing.* Joint Statistical Meetings Virtual Conference. <https://ww2.amstat.org/meetings/proceedings/2021/data/assets/pdf/1913735.pdf>

Keller, A., Mule, T., & Barth, D. (2024). *Predicting Vacant Housing Units in the American Community Survey.* U.S. Census Bureau. https://www.census.gov/library/working-papers/2024/acs/2024_Keller_01.html



Thank You!

Dorothy.A.Barth@census.gov