

# Using Administrative Data in the American Community Survey

Elizabeth Poehler, U.S. Census Bureau  
2023 ACS Data Users Conference  
May 16, 2023

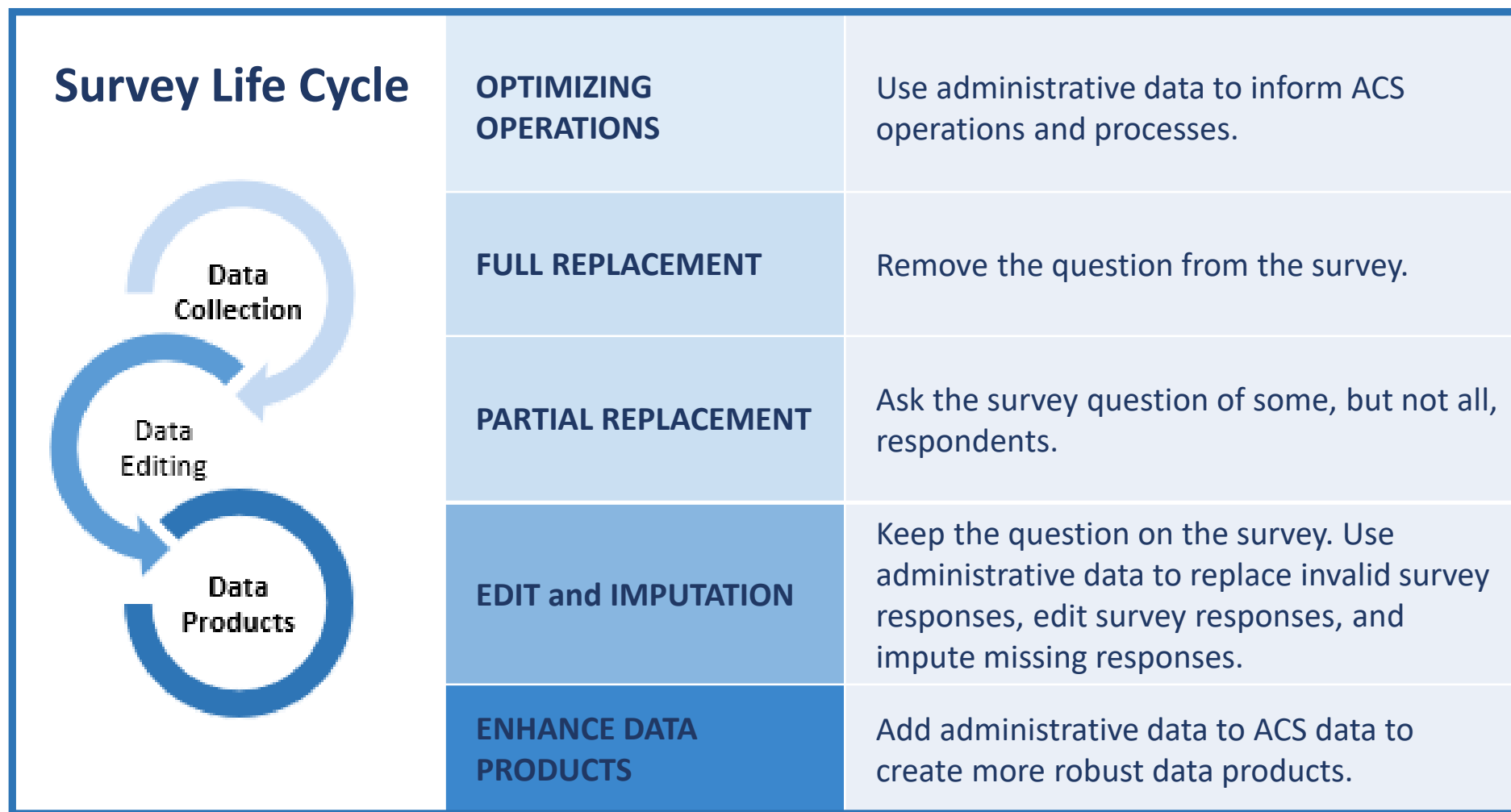
# Why Use Administrative Data?

- ▶ Increasing push back from the public to respond to the survey
- ▶ Increasing concerns about the potential for nonresponse bias in our data
- ▶ Title 13 requires that we consider data already available before collecting it from the public
- ▶ Reduce respondent burden and produce high quality data

# What is Administrative Data?

- ▶ Data from sources such as:
  - federal agencies
  - state and local governments
  - commercial entities
- ▶ These sources may maintain data for the purpose of administering programs, providing services to the public, or for business use.

# Administrative Data Uses in the ACS



# OPTIMIZING OPERATIONS

# Why do we need to optimize CAPI operations?

- ▶ CAPI = nonresponse follow-up operation for people who don't self-respond to the survey.
  - In person and telephone interviewing is conducted during CAPI.
- ▶ Interviewing is more expensive now than it has ever been
  - Increase in wages and overhead costs
  - Decrease in respondent cooperation, means more visits to complete an interview

# FY22 CAPI Cost Saving Strategies

- Reduced CAPI workload
  - October – March: typical workload
  - April – June: reduced by a several thousand cases
  - August: cut workload in half
  - September: back to typical workload
- Reduced the time allowed to conduct interviews
  - October – June: 1 month to conduct interviews
  - July – September: reduced time by 2 days to 2 weeks

# Optimizing CAPI Data Collection

1. ***Prioritizing*** Contacts by Geographic Areas
2. ***Predicting*** Vacancy Status
3. ***Adapting*** Contact Efforts



# Prioritizing Contacts by Geographic Areas

**We are prioritizing contacts by geographic areas for the CAPI workload based on research of:**

- ▶ Response rates
- ▶ Administrative data
- ▶ Respondent demographic characteristics

# Predicting Vacancy Status

- ▶ We created a model to predict which housing units have a high probability of being vacant.
- ▶ The model uses data from:
  - ▶ U.S. Postal Service
  - ▶ Census Bureau's Master Address File
  - ▶ IRS
  - ▶ Medicare

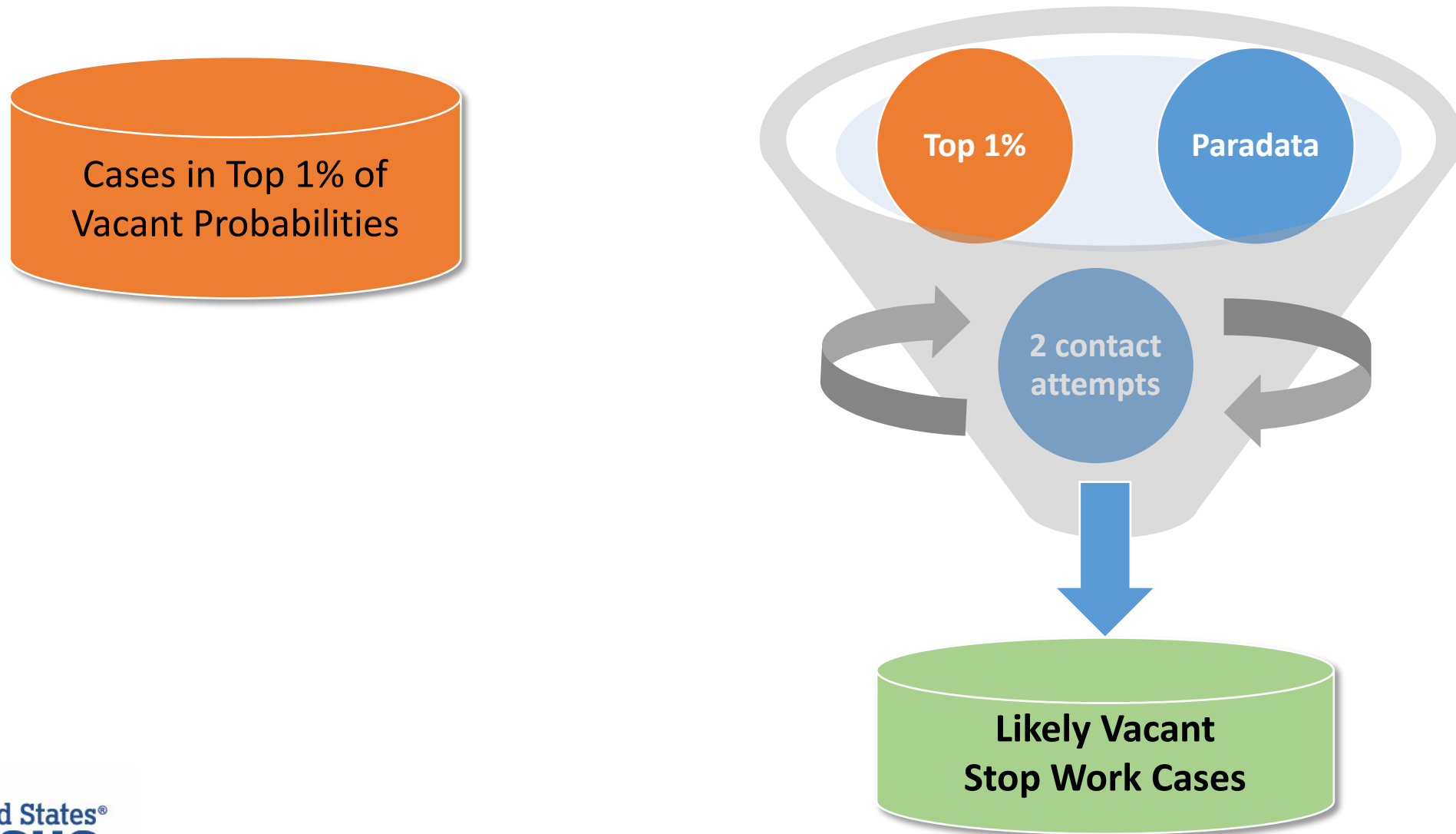
# Predicting Vacancy Status

Cases in Top 3% of  
Vacant Probabilities

Personal Visit  
First



# Predicting Vacancy Status



# Adapting Contact Efforts

- ▶ Complements the Other Two Projects
  - Focuses modeling on likely occupied housing units
  - Includes a case-by-case data-driven approach to adapt contact efforts
- ▶ Borrows from models and strategies currently used in other surveys
  - SIPP, NSCG, NSFG, Labor Force Surveys @ Statistics Netherlands
  - Demonstrated effort redirection and cost reduction benefits in other surveys

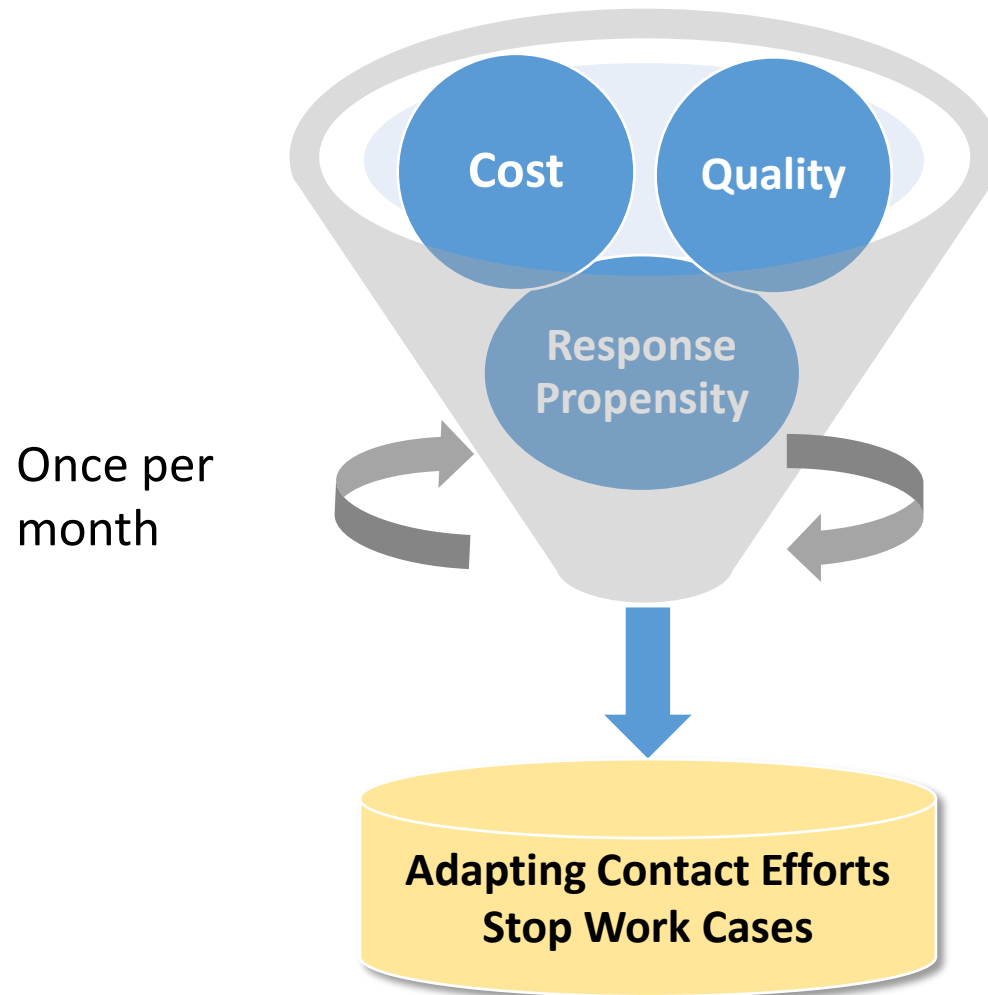
# Adapting Contact Efforts

**Response Propensity:** Case-level response propensity models

**Quality:** Responding and non-responding housing unit characteristics

**Cost:** Case-level data collection cost models

# Adapting Contact Efforts



# FULL AND PARTIAL QUESTION REPLACEMENT



# Partially Replace the Acreage Question

- ▶ Use Property Tax and Parcel Boundary Data to determine acreage
- ▶ Evaluate Quality
  - Compared administrative data and ACS data at county level
- ▶ Establish Business Rules:
  - Internet and CAPI instruments skip acreage
  - Paper: ask acreage - edit process will select value
- ▶ 2024 ACS Implementation

## 4 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

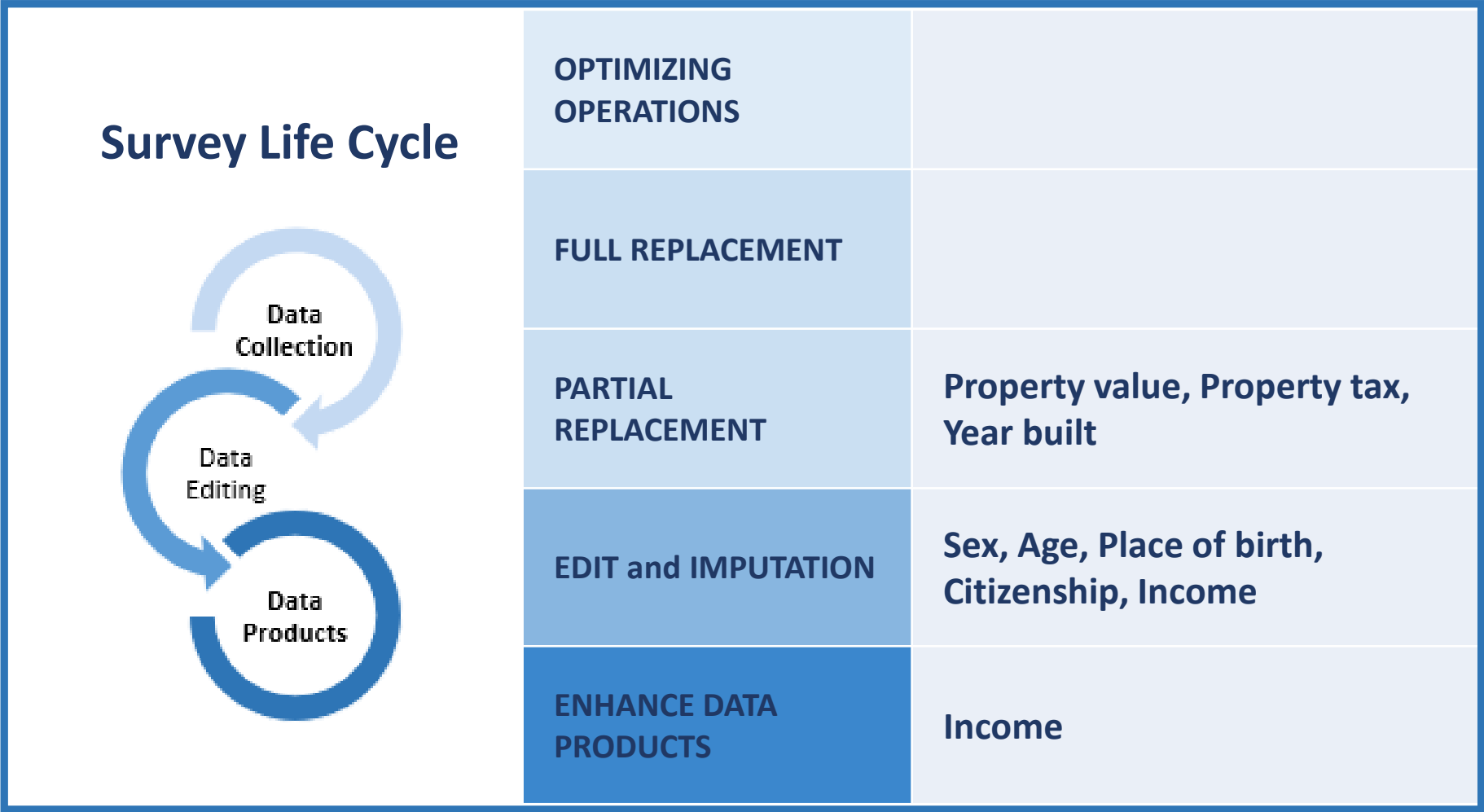
# Potentially Remove Agricultural Sales Question

- ▶ Use IRS tax data and crop-layer data to identify farms
- ▶ Evaluate Quality
  - Simulate estimates (joint effort with BEA)
  - Research findings anticipated winter 2023
- ▶ Make Recommendation
  - Potential 2025 ACS Implementation

5

**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



# Conclusion

- ▶ The Census Bureau has a shared goal with our data users of high quality statistical products.
- ▶ The Census Bureau also has a responsibility to the public to use existing data before burdening the public with requests for information.
- ▶ We are committed to working with the public and our data user community to balance these demands.
- ▶ We believe that the use of administrative data can enhance survey operations and the products we produce.