# Data Reconciliation and Estimation in an Agricultural Survey

Habtamu Benecha
Denise A. Abreu
Rachael Jennings
Linda J. Young

National Agricultural Statistics Service (NASS)
United States Department of Agriculture

Joint Statistical Meetings
August 4, 2020

### Disclaimer

The findings and conclusions in this presentation are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

# Background

- ▶ NASS conducts the Puerto Rico Census of Agriculture
  - ▶ A complete count of Puerto Rico farms and the people who operate them
- ▶ Farm - at least $500 in agricultural products
- ▶ Census Mailing List (CML) - sampling frame for the census
  - ▶ A list of farms and potential farms in Puerto Rico
- ▶ The CML does not cover all farms - undercoverage
- ▶ The Agricultural Coverage Evaluation Survey (ACES)
- ▶ Census estimation

# Background

- The Puerto Rico Census was to be conducted in 2017
- Data were collected for the 2017 ACES
- Then came Hurricane Maria in September 2017
  - Many farms destroyed
  - Infrastructure damaged
- The 2017 Census was delayed by one year, conducted in 2018
- The ACES was repeated in 2018 using the same sample

An area in PR before (left) and after (right) Hurricane Maria

# Challenges and Goals

- Data collection procedures were different in the two years
- Difficulty to apply record linkage

**Goals:**

- Reviewing the 2017 and 2018 ACES records
- Estimating the number of farms and land in farms before and after Hurricane Maria

# The Agricultural Coverage Evaluation Survey (ACES)

- To provide a measure of undercoverage of the CML
- Based on an area frame, which covers all land in Puerto Rico
- Primary sampling units (PSU) are drawn using stratified random sampling
- A segment of land is randomly selected from each PSU

# The Agricultural Coverage Evaluation Survey (ACES)

Tracts in a segment

# Review of the 2017 and 2018 ACES records

- ▶ The 2017 and 2018 ACES data were collected from the same segments
- ▶ Not all tract delineations from 2017 were available for 2018
- ▶ So, the same tract number may represent different tracts in the 2017 and 2018 surveys
- ▶ Automated record linkage not possible

# FEITH images

- Images of completed questionnaires (FEITH images) available for both 2017 and 2018 ACES

# Comparing address and name information; 2017 and 2018 ACES

- The FEITH images contain addresses and operator names
- Manually compared names/addresses on the images for all tracts in each segment
- Linked many of the 2017 and 2018 tracts
- Some of the tracts could not be linked
  - No data available for a tract (i.e., No-Link)
  - Differences in names and addresses (i.e., Uncertain-Link)

# Census Estimation: Capture-recapture approach

- Capture-recapture method
- Two sources of data
  - Census
  - ACES
- Probabilistic record linkage
- Independent sources of data
- Census weights
  - Undercoverage
  - Non-response
  - Misclassification
    - Classifying farms as non-farms
    - Classifying non-farms as-farms

# ACES Estimation

- ACES - can be used for the estimation of the number of farms, land in farms and other commodities
- A stratified random sample drawn from an area-frame
- Final weights are estimated using an expression of the form

$$W_j = b_j c_j, \tag{1}$$

Where, $b_j$ is the sampling weight for tract $j$, and $c_j$ is the tract-to-farm ratio

# Manual Record Linkage Results

Table: 2017 and 2018 Record Linkage Results: Numbers of Records

|                      | Year 2017 | Year 2018 |
|----------------------|-----------|-----------|
| Linked Farms         | 219       | 198       |
| Not Linked Farms     | 72        | 134       |
| Linked Non-Farms     | 702       | 723       |
| Not Linked Non-Farms | 338       | 2,115     |
| Total                | 1,331     | 3,170     |

# Manual Record Linkage Results

- ▶ Tracts that were not linked
  - ▶ No-link
  - ▶ Uncertain-link
- ▶ 'Not-Linked' tracts - all tracts with no links or uncertain links

Table: 2017 and 2018 - Numbers of Farm Records Not Linked

|                | Year 2017 | Year 2018 |
|----------------|-----------|-----------|
| No link        | 31        | 95        |
| Uncertain link | 41        | 39        |
| Total          | 72        | 134       |

# Manual Record Linkage Results

Table: Year of Establishment for 2018 Farm Records Not Linked

|  |  | Count |
|---|---|---|
| Farms Prior to 2017 | No-Link | 87 |
|  | Uncertain Link | 35 |
| New Farms Since 2017 | No-Link | 8 |
|  | Uncertain Link | 4 |
| Total |  | 134 |

# Adjustment Estimation

- Unadjusted estimates were produced using original ACES data
- Tracts with no links and uncertain links were considered to produce adjusted estimates
- Two sets of adjusted estimates
  - Tracts with no links were added to the original ACES data to produce the first set of adjusted estimates
  - Tracts that were not linked (i.e., tracts with no links and uncertain links) were added to the original ACES data to produce a second set of adjusted estimates
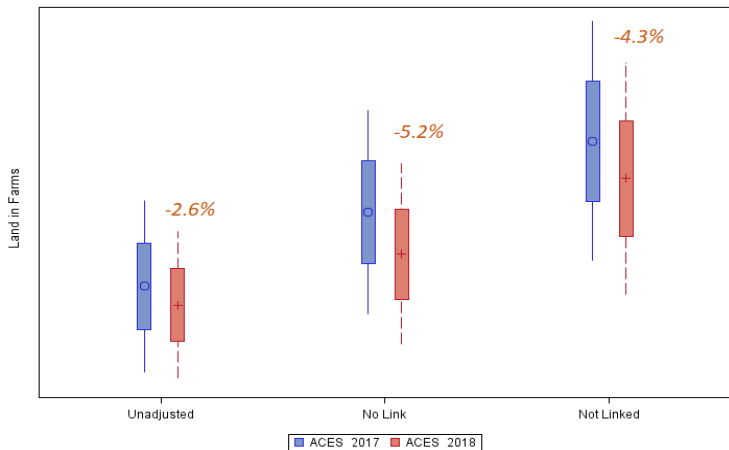- Not all tracts with no links or uncertain links may be operational

# Number of Farms Estimates: 2017 and 2018 ACES
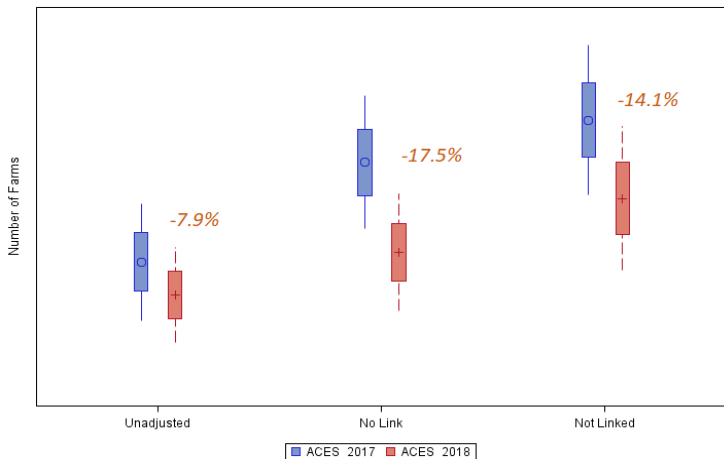
# Number of Farms Estimates: ACES 2018 vs Census 2018
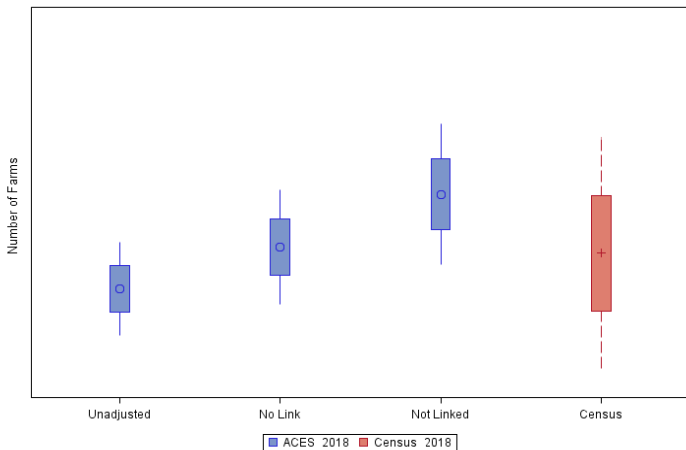
# Land in Farms Estimates: 2017 and 2018 ACES

# Land in Farms Estimates: ACES 2018 vs Census 2018
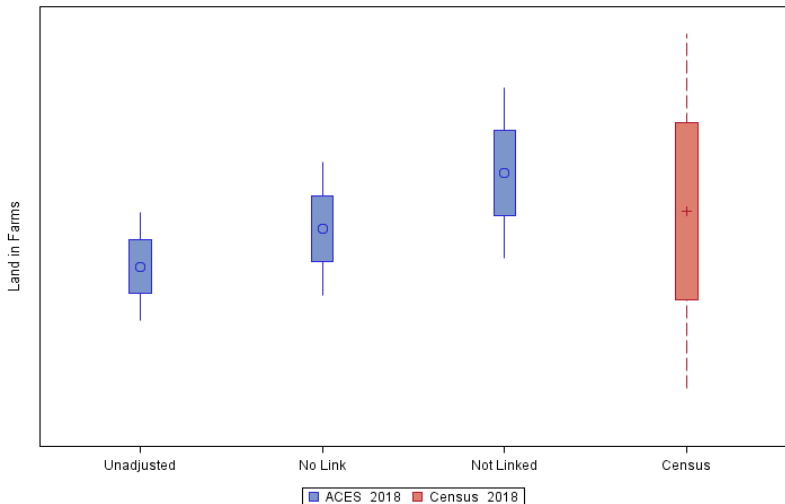
# Summary

- Number of farms and land in farms before and after Hurricane Maria
- Both the unadjusted and adjusted estimates showed overall decreases in the number of farms and land in farms from 2017 to 2018
- The number of farms established after Hurricane Maria
- Farms that survived the hurricane but not captured in the ACES
- ACES helped validate the 2018 Census results
- Future work: applying a capture-recapture approach for estimation from the ACES

# References

Abreu, D.A., J.S. McCarthy, L.A. Colburn (2010). Impact of the Screening Procedures of the June Area Survey on the Number of Farms Estimates. Research and Development Division. RDD Research Report Number RDD-10-03. Washington, DC: USDA, National Agricultural Statistics Service.

Young, Linda J., Andrea C. Lamas, and Denise A. Abreu (2017). The 2012 Census of Agriculture: a capturerecapture analysis. Journal of Agricultural, Biological and Environmental Statistics 22, no. 4: 523-539.

Young, Linda J., Andrea C. Lamas, Denise A. Abreu, Shu Wang, and Daniel Adrian (2013). "Statistical methodology for the 2012 us census of agriculture." In the Proceeding 59th ISI World Statistics Congress, pp. 1063-1068.

Lamas, Andrea C., Denise A. Abreu, Pam Arroway, Kenneth K. Lopiano, and Linda J. Young (2010) Modeling misclassification in the June Area Survey. Proceedings of the 2010 Joint Statistical Meetings, ASA Section on Survey Research Methods 24802488.

Benecha, HK, Denise A, Abernethy J, Sartore L, Young L. (2017). Evaluation of a New Approach for Estimating the Number of U.S. Farms. In JSM Proceedings, Survey Research Methods Section. Baltimore, MD: American Statistical Association

# References

Abreu, D. (2007). Results from the 2002 classification error study. Research and Development Division. RDD Research Report: RDD-07-03.

Abreu, D., Arroway, P., Lamas, A., Lopiano, K., and Young, L. (2010a). Using the census of agriculture list frame to assess misclassification in the June area survey. In Proceedings of the Joint Statistical Meetings.

Abreu, D., Busselberg, S., Lamas, A., Barboza, W., and Young, L. (2014). Evaluating a new approach for estimating the number of U.S. farms with adjustment for misclassification. In Proceedings of the Joint Statistical Meetings.

Abreu, D., Dickey, N., and McCarthy, J. (2009). 2007 classification error survey for the united states census of agriculture. Technical report, United States Department of Agriculture, National Agricultural Statistics Service.

Abreu, D., McCarthy, J., and Colburn, L. (2010b). Impact of the screening procedures of the June Area Survey on the number of farms estimates. Research and Development Division. RDD Research Report RDD-10-03. Washington, DC: USDA, National Agricultural Statistics Service.

Lopiano, K., Lamas, A., Abreu, D., Arroway, P., and Young, L. (2011). Adjusting the June Area Survey estimate of the number of U.S. farms for misclassification and non-response. Technical report, United States Department of Agriculture, National Agricultural Statistics Service.

Young, L., Abreu, D., Arroway, P., Lamas, A., and Lopiano, K. (2010). Precise estimates of the number of farms in the united states. In Proceedings of the Joint Statistical Meetings.

# Thank You!

Contact:

`habtamu.benecha@usda.gov`