Successes and Challenges in Assessing Data Quality of Non-Survey Data Sources: Impacts on Scientific Integrity

Denise A. Abreu, Luca Sartore,
Linda J. Young
National Agricultural Statistics Service (NASS)

October 23, 2024

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.

Scientific Integrity

- The adherence to professional practices, ethical behavior, and the principles of honesty and objectivity when conducting, managing, using the results of, and communicating about science and scientific activities. Inclusivity, transparency, and protection from inappropriate influence are hallmarks of scientific integrity

- *New* Federal government-wide definition
  - Source: A Framework for Federal Scientific Integrity Policy and Practice (whitehouse.gov)

USDA Scientific Integrity Policy (SIP)

U.S. DEPARTMENT OF AGRICULTURE
WASHINGTON, D.C. 20250

DEPARTMENTAL REGULATION	NUMBER: DR 1074-001
SUBJECT: Scientific Integrity	DATE: May 2, 2024
OPI: Office of the Chief Scientist	EXPIRATION DATE: May 2, 2029

Departmental Manual, DM 1074-001

PROCEDURES FOR RESPONDING TO ALLEGATIONS OF COMPROMISED SCIENTIFIC INTEGRITY

- For the purposes of this regulation:
  - (a) Scientific Integrity is also the condition resulting from such adherence
  - (b) This condition ensures objectivity, clarity, and reproducibility, while providing insulation from fabrication, falsification, plagiarism, bias, inappropriate influence, political interference, censorship, as well as from inadequate procedural and information security

Scientific Integrity Principles:
A real-life application – June Area Survey (JAS)

- NASS's largest annual survey
- Conducted via face-to-face interviews since the 1950s
- Annual estimate of the number of farms and land in farms in US
- Direct estimates of large crop commodities such as corn, soybeans, wheat, etc.
- Measures the incompleteness of the NASS list frame

COVID-19 Challenges the Scientific Integrity of the JAS

- **March 2020** – Inability to conduct in-person interviews led to cancellation of 2020 JAS
- **October 2020** – Leadership decided to conduct 2021 JAS without in-person interviews
- **Fall 2021** – Leadership decided 2021 changes would remain for 2022 JAS, with limited in-person interviews allowed

COVID-19 Challenges the Scientific Integrity of the JAS

- Response rates have been on the decline for all surveys, regardless of mode (Czajka et al., 2016; Johansson et al., 2017)

- Literature shows that face-to-face interviews provide better quality data (Heerwegh et al., 2008; Blumberg et al., 2021)

Task

What scientific integrity principles/practices can be applied to safeguard the reliability of the JAS under this new data collection paradigm?

NASS -- June Area Survey (JAS)

- Area-frame based
- Segments of land sampled
- Sampled segments divided into tracts representing unique land operating arrangements

NASS -- June Area Survey (JAS)

- Rotating panel design -- 20% of the sample enters each year and remains for 5 years

- Each yearly sample
  - 20% New segments
  - 80% Old segments

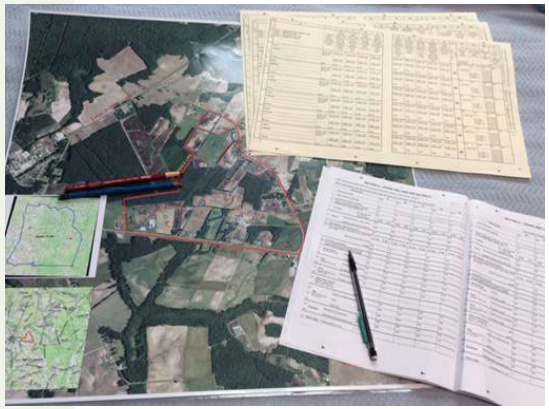
Traditional JAS Data Collection Timeline

Prescreening
May

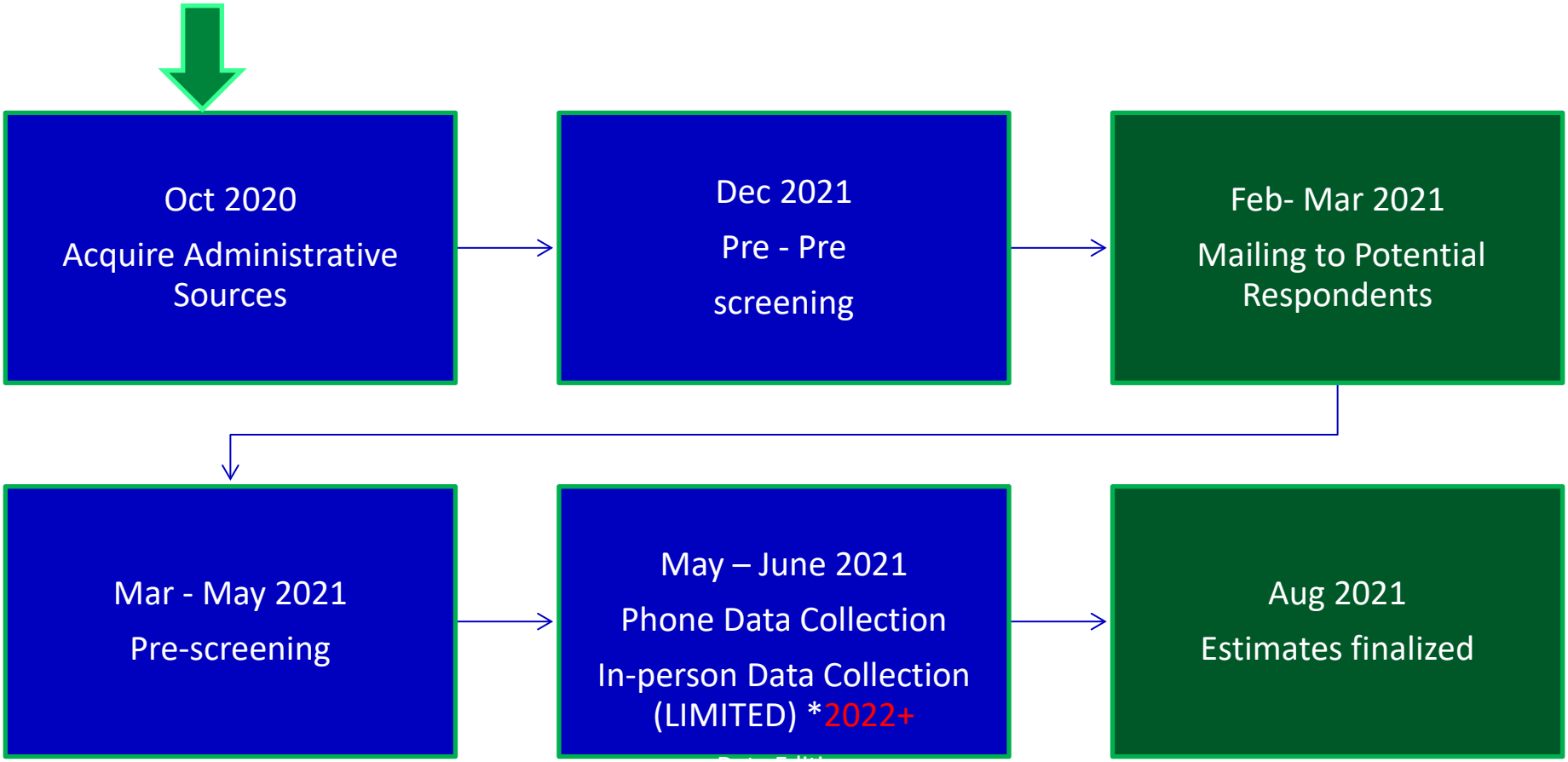
- Focus on new segments
- "Blank" 24"x24" map
- Listing of potential operators
- Prescreening names
- Identify non-agricultural tracts

Data Collection
June

- First two weeks of June
- Return to agricultural tracts
- Collect tract and whole farm information
- Hand impute refusals and inaccessibles



JAS Mixed Mode Data Collection Timeline

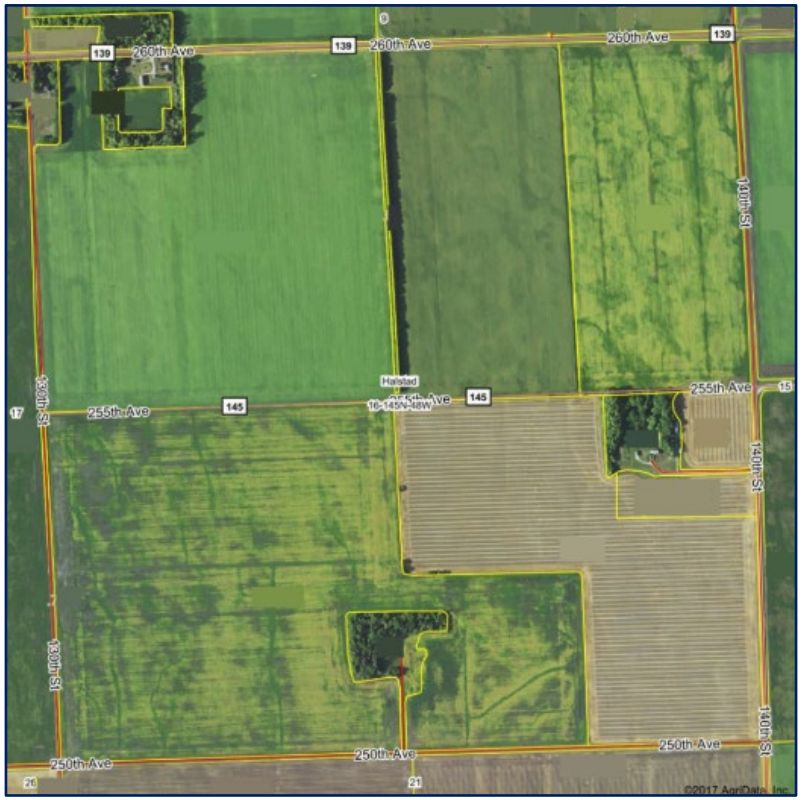


Non-Survey Data Sources

- Quality assessment of non-survey sources
- Usefulness of non-survey data
- Rigorous and comprehensive evaluation

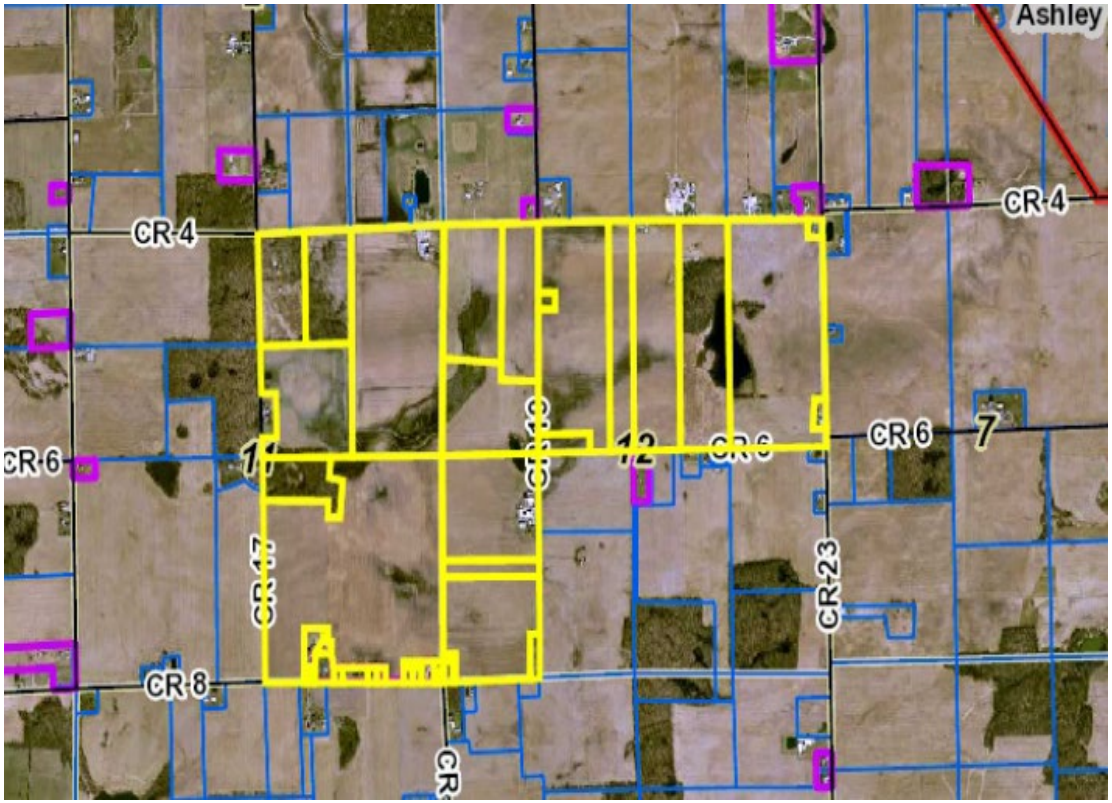
Administrative Data – Farm Service Agency (FSA)

- FSA Common Land Units (CLUs)
  - Geospatially referenced boundaries
  - Basically, correspond to a field (single crop, same operator, etc.)
- Form: FSA-578
  - Crop information for growing season
  - Gold standard for crops planted



FSA Common Land Unit and 578 Data

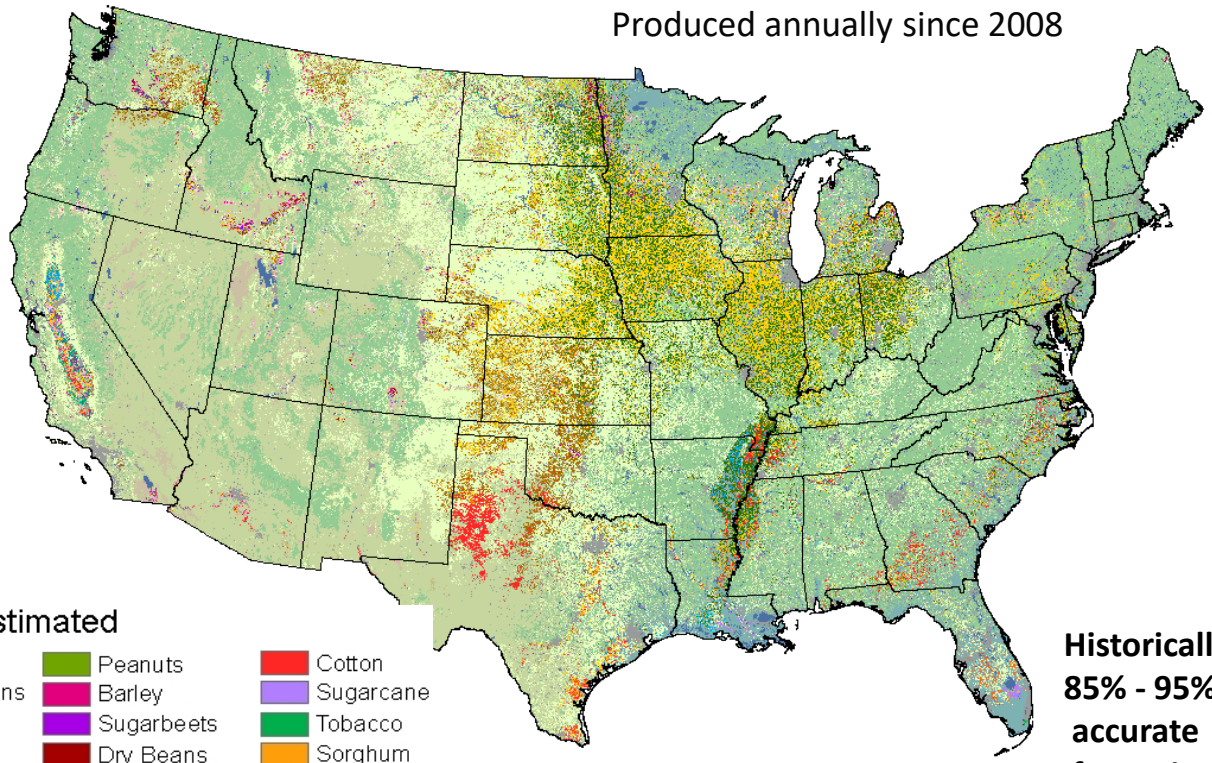
County Tax Assessors Parcel Information



- Administrative boundaries maintained by county tax assessor offices
- Contain basic contact information associated with land ownership
- Free to the public but disaggregated across county administrations
- NASS partnered with vendor, CoreLogic, to obtain nationwide dataset

Cropland Data Layer (CDL)

Georeferenced, crop specific, land cover data set
Raster product at 30-meter resolution
Produced annually since 2008



Crops Estimated

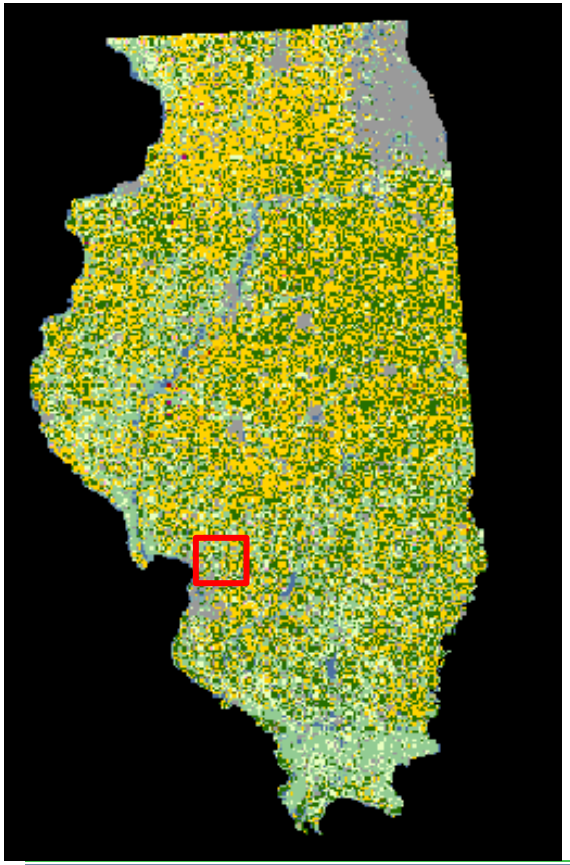
Corn	Peanuts	Cotton
Soybeans	Barley	Sugarcane
Alfalfa	Sugarbeets	Tobacco
Rice	Dry Beans	Sorghum
Canola	Spring Wheat	Potatoes
Flaxseed	Winter Wheat	
Sunflower	Durum Wheat	

*** 9 billion pixels 30m product**

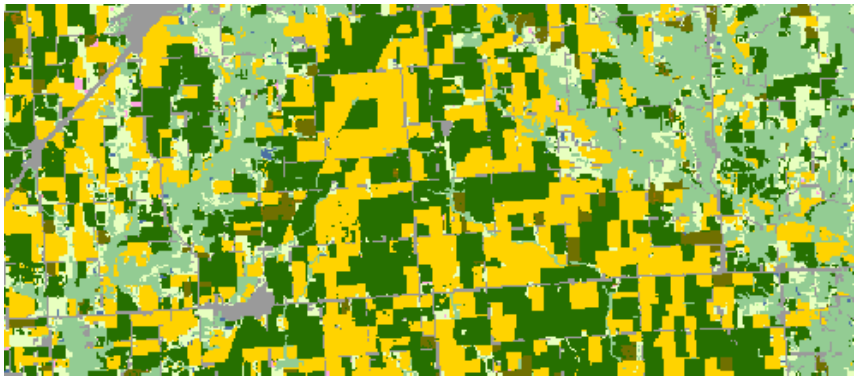
Historically, 85% - 95% accurate for major crops

Predictive Cropland Data Layer (PCDL)

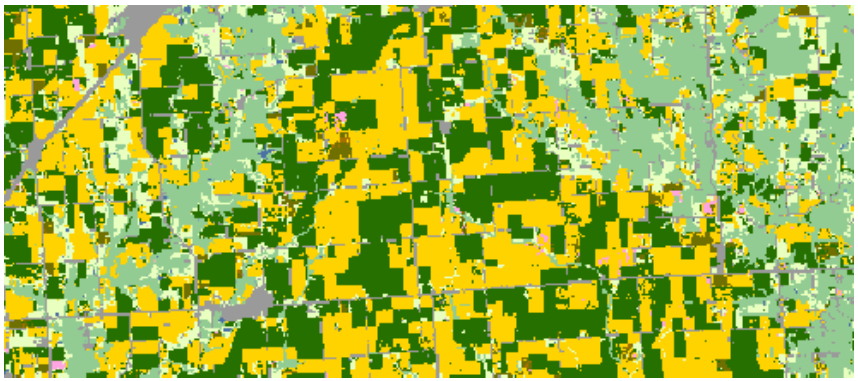
2018 Predictive CDL for Illinois



2018 Predictive CDL



2018 – Final CDL

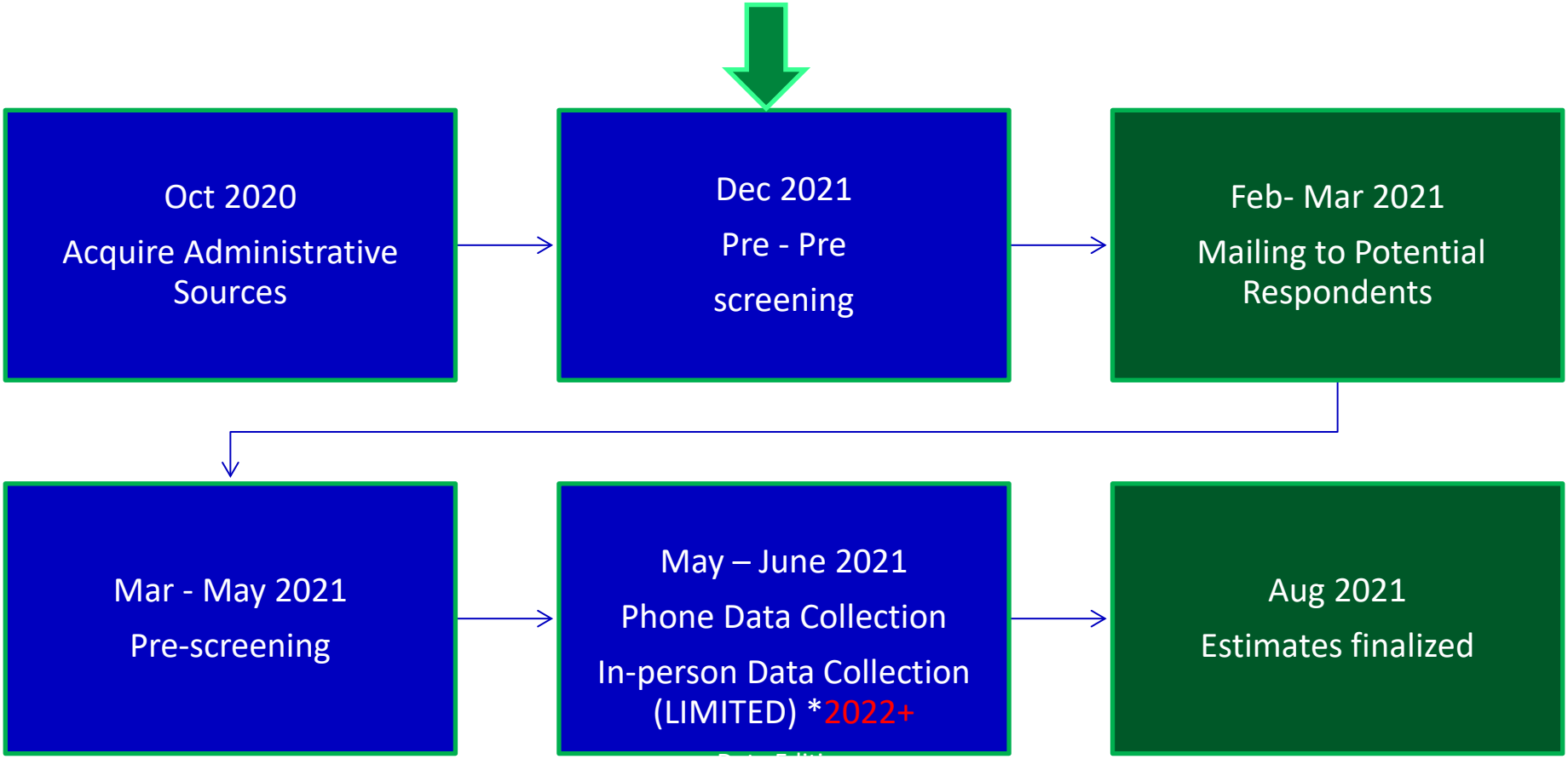


- Predicts crops at the field level
- Not identical but close to actual planted acres
- About 80 percent accurate

USDA

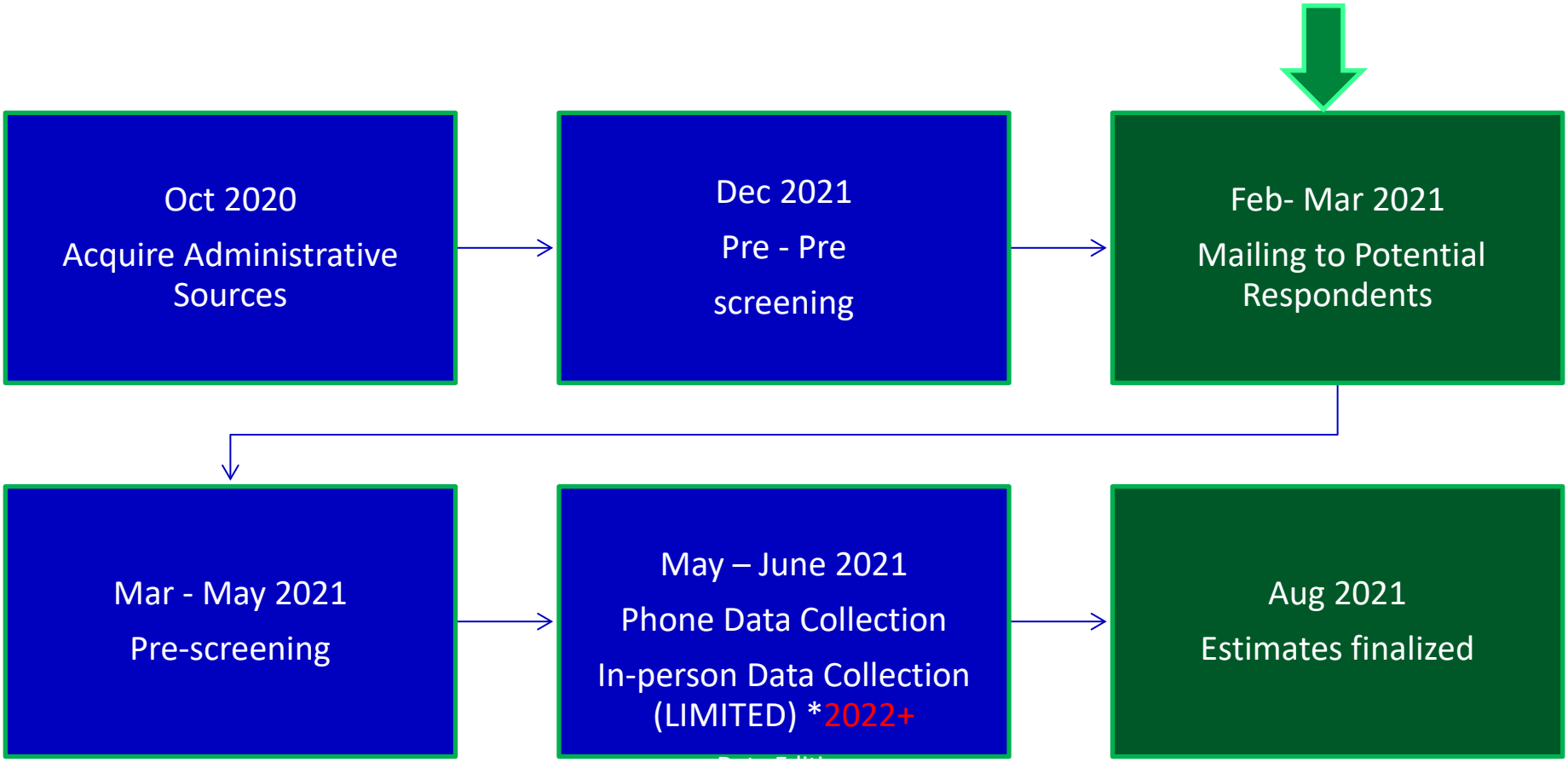
AGRICULTURE COUNTS

JAS Mixed Mode Data Collection Timeline



- Reproducibility of the data collection strategy
- Robust and appropriate design for identifying likely respondents in the population of interest
- Strengthening the process through feedback loops

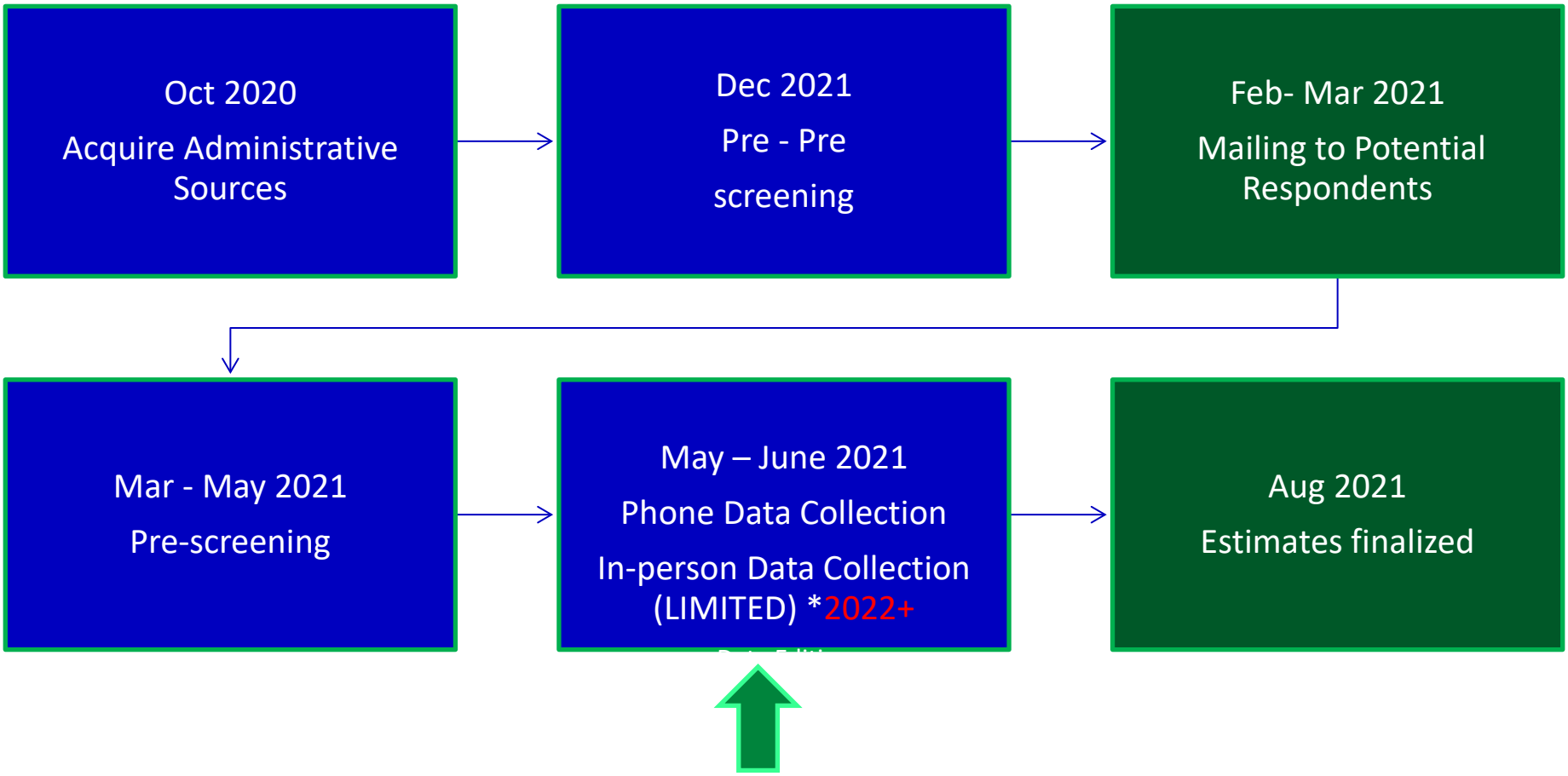
JAS Mixed Mode Data Collection Timeline



Segment Maps Mailed

- Data collection strategy designed to maintain high quality standards
- Bias reduction and uncertainty control
- Quality assessments on final estimates

JAS Mixed Mode Data Collection Timeline



June Area Land Tool

- Training personnel on the proper usage of novel tools
- Oversight of data acquisition and editing processes by qualified personnel

Extensive Research and Evaluations on Quality of JAS Data

- Peer review and results dissemination (including unanticipated findings)
- Transparency based on accepted standards under given regulations on privacy and confidentiality

References

- A Framework for Federal Scientific Integrity Policy and Practice (whitehouse.gov)
- USDA Office of the Chief Scientist Website https://www.usda.gov/our-agency/staff-offices/office-chief-scientist-ocs
- Departmental Regulation 1074-001 (“Scientific Integrity”) https://www.usda.gov/directives/dr-1074-001
- Czajka, J. L., and Beyler, A. (2016). Background Paper Declining Response Rates in Federal Surveys: Trends and Implications. Mathematica Policy Research 1: 1-86. Available at: Declining Response Rates in Federal Surveys: Trends and Implications (hhs.gov)
- Johansson, R., Effland, A., and Coble, K. (2017). Falling Response Rates to USDA Crop Surveys: Why it Matters. Farmdoc Daily 7(9). Available at: Falling Response Rates to USDA Crop Surveys: Why It Matters - farmdoc daily (illinois.edu)
- Dirk Heerwegh, Geert Loosveldt, Face-to-Face versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality, Public Opinion Quarterly, Volume 72, Issue 5, December 2008, Pages 836–846, https://doi.org/10.1093/poq/nfn045
- Stephen J. Blumberg, Jennifer D. Parker, and Brian C. Moyer, 2021:National Health Interview Survey, COVID-19, and Online Data Collection Platforms: Adaptations, Tradeoffs, and New Directions American Journal of Public Health 111, 2167_2175, https://doi.org/10.2105/AJPH.2021.306516
- Abreu, D. A., Murphy, T., Rosales, A., Young, L. J. (2024) “From In-Person to Mixed Mode: Trends and Characteristics of Respondents to the June Area Survey.” Presented at the Seventh International Conference on Establishment Statistics, Glasgow, UK.
- Murphy, T., Rosales, A., Abreu, D. A., Young, L. J. (2024) “Resilience of an Area Survey Program in the Face of Change.” Presented at the Seventh International Conference on Establishment Statistics, Glasgow, UK.
- Murphy, T., Rosales, A., Abreu, D. A., Young, L. J. (2023) “Impact of Shifting Data Collection Mode on Area Survey Data Quality.” Presented at the Ninth International Conference on Agricultural Statistics, Washington, DC.

Thank you for your attention!!

Denise A. Abreu

Denise.Abreu@usda.gov