# Analyzing and Using Web Paradata to Understand and Improve the Data Collection Process

Robyn Sirkis and Pam McGovern

National Agricultural Statistics Service

FedCASIC 2019

# Outline

- Background

- Paradata Results

- Recommendations and Next Steps

# 2017 Census of Agriculture

- A complete count, taken every five years, of America's farms and ranches and the people who operate them.

- The only source of uniform, comprehensive and impartial agricultural data for every county in the nation.

- Questionnaire contains 34 sections collecting information on land use and ownership, operator characteristics and demographics, production practices, income and expenditures, and more.

- Data Collected: from November 24,2017 to August 29,2018

# Extracting the Web Paradata

- Complex files contained the following information:
  - types of devices and browsers
  - date and time of each web instrument session
  - the value of each question answered
  - Warning screens triggered/clicked help button/clicked save button
- Examples of challenges:
  - Obtaining device and browser

> **User agent String (device=tablet, browser=internet explorer):** Mozilla/5.0 (Windows NT 6.3; WOW64; Trident/7.0; Touch; .NET4.0E; .NET4.0C; .NET CLR 3.5.30729; .NET CLR 2.0.50727; .NET CLR 3.0.30729; Tablet PC 2.0; F9J; CMNTDFJS; InfoPath.3; rv:11.0) like Gecko

# Extracting the Web Paradata Continued

- Extracting location where warning screens triggered/clicked help button/clicked save button
  - Save button (choose question before), help button (choose question after)

- Calculating change of answers metrics
  - Respondent can backspace and re-enter same answer or can switch back and forth between two answers (for example)

- Determining location of break-offs
  - Need to determine furthest question in the form completed since the respondent can go back to previous questions

# Rates and Device Usage

- How many respondents are submitting the web form once they start it?
  - Number Submitted: 449,604
  - Return Rate: 24.1%
  - Cooperation Rate: 96.3%
  - Break-off Rate: 3.7%

| Cooperation and Break-Off Rates by Device | | | |
|---|---|---|---|
| **Device** | **Percent by Type** | **Number Accessed Questionnaire** | **Cooperation Rate** |
| Desktop | 83.6 | 389,352 | 96.5 |
| Tablet | 10.0 | 46,387 | 96.1 |
| Phone | 6.1 | 28,369 | 94.0 |
| E-Book Reader | 0.4 | 1,729 | 96.4 |

# Mobile Devices

- What types of mobile devices are used by the respondent?

| Mobile Device Type | |
|---|---|
| **Device** | **Percent** |
| iPad | 37.0 |
| Other Tablet | 23.7 |
| iPhone | 20.5 |
| Android Phone | 12.5 |
| Samsung Phone | 4.0 |
| E-Book Reader | 2.3 |

| Orientation for Mobile Devices | | |
|---|---|---|
| **Orientation** | **Phone** | **Tablet** |
| **Landscape** | 9.1 | 75.5 |
| **Portrait** | 90.9 | 24.4 |

# Sessions Information

- Do respondents submit the form in one or more sessions?
  - 19.6% : multi-session users
  - Average session time approximately 28.7 minutes

| Number of Sessions | |
|---|---|
| **Sessions** | **Percent** |
| 1 | 80.4 |
| 2 | 13.2 |
| 3 | 3.8 |
| 4 | 1.4 |
| 5+ | 1.3 |

| Sessions By Device | | | | |
|---|---|---|---|---|
| **Number of Sessions** | **Desktop** | **Phone** | **Tablet** | **E-Book Reader** |
| 1 | 80.2 | 79.7 | 84.4 | 78.7 |
| 2 | 13.4 | 13.1 | 10.4 | 13.7 |
| 3 | 3.9 | 3.9 | 2.9 | 4.0 |
| 4 | 1.4 | 1.6 | 1.1 | 1.7 |
| 5 | 0.6 | 0.7 | 0.4 | 0.7 |
| 6+ | 0.6 | 1.1 | 0.8 | 1.2 |

# Switching Devices

- Do respondents switch devices when completing the form in more than one session?
  - 5.3% of the multi-session users switched devices

| Switching Devices  (Top 5) | | |
|:---:|:---:|:---:|
| **First Device** | **Last Device** | **Percent** |
| Tablet | Desktop | 34.3 |
| Phone | Desktop | 21.6 |
| Desktop | Tablet | 21.0 |
| Desktop | Phone | 6.5 |
| Desktop | Desktop | 4.7 |

# Browser Type

- What browser is the respondent using on their device?

| Browser Type | |
|---|---|
| **Browser** | **Percent** |
| Chrome | 40.9 |
| Internet Explorer | 17.9 |
| Safari | 16.3 |
| Edge | 13.4 |
| Firefox | 10.2 |
| Samsung Browser | 0.6 |
| Silk | 0.4 |
| Opera | 0.2 |
| Other | 0.1 |

# Day of the Week and Hours

- What day and time are most respondents submitting the form?
  - Highest percentages were Monday through Thursday
  - Highest percentages mid-morning and early afternoon
    - 10am to 10:59am: 8.9%
    - 7am to 7:59am: 3.3%
    - 9pm to 9:59pm: 3.2%

| Day of the Week | |
|---|---|
| **Day of Week** | **Percent** |
| Monday | 18.4 |
| Tuesday | 17.0 |
| Wednesday | 15.4 |
| Thursday | 14.1 |
| Friday | 12.4 |
| Sunday | 12.2 |
| Saturday | 10.5 |

# Warning Messages

- What warning screens are respondents triggering the most?
  - 11 warning messages
  - 41.0% respondents triggered warning messages
    - Respondents need to get total acres to match between two sections
    - Approximately 37% of the respondents that triggered the email warning message left the email address question blank

| | | Warning Messages (Top Two) | |
|---|---|---|---|
| **Section** | **Section Name** | **Warning Messages** | **Percent of Respondents** |
| 2 | Land Use in 2017 | The total land use acres do not equal the total acres operated reported in the previous section on Acreage. | 34.2 |
| 34 | Conclusion | Please enter a valid email address. | 34.0 |

# Help and Save Buttons

- How many respondents are selecting the help and save buttons?
  - 4.7% respondents selected help
    - Majority select help at the beginning of the questionnaire
    - Respondents can print the help document
  - 20.7% respondents select save
    - Majority select save at the end of the questionnaire

| Help Screens (Top Two) | | |
|---|---|---|
| **Section** | **Section Name** | **Percent of Respondents** |
| 0 | Out-of-Business Screeners | 22.0 |
| 1 | Acreage in 2017 | 17.7 |

# Break Offs

- Where are respondents breaking off in the form?

| Break Offs by Section | | |
|---|---|---|
| Section Number | Section Name | Percent |
| 30 | Production Expenses | 23.5 |
| 1 | Acreage in 2017 | 11.3 |
| 0 | Out-of-Business Screeners | 10.0 |
| 2 | Land Use in 2017 | 9.3 |
| 7 | Personal Characteristics | 8.7 |

# Changing Answers

- Which questions are respondents changing their answers the most?
  - Aquaculture: identified issues with exclude/include statements
  - More respondents changed answers to acres questions as opposed to yes/no/categorical questions

| | Changing Answers by Question | | |
|---|---|---|---|
| **Section Number** | **Section Name** | **Question Description** | **Percent** |
| 18 | Aquaculture | Aquaculture specify write-in | 36.0 |
| 2 | Land Use in 2017 | Permanent Pasture and Rangeland, Acres | 20.6 |
| 14 | Cattle and Calves | Beef Cow--Inventory | 18.8 |
| 2 | Land Use in 2017 | Cropland Harvested, Acres | 18.4 |
| 2 | Land Use in 2017 | All Other Land, Acres | 18.2 |

# Changing Answers to Screener Questions

- How many respondents change their answers to screener questions from yes to no?
  - Examined 24 screener questions

| Changing Answers from Yes to No | | |
|---|---|---|
| **Section Name** | **Question Description** | **Percent** |
| Food Marketing Practices | Produce & Sell Crops/Livestock/Poultry or Ag Products for Human Consumption | 7.9 |
| Conclusion | Is it possible that the agricultural activity information reported for this operation could also be reported by another operation? | 3.8 |
| Fertilizers, Chemicals and Soil Conditioners Applied | Any Fertilizer or Chemicals | 3.7 |
| Conclusion | Any Other Farm | 2.1 |

# Demographic Analyses

- Are their differences in demographics across device?



Age by Device
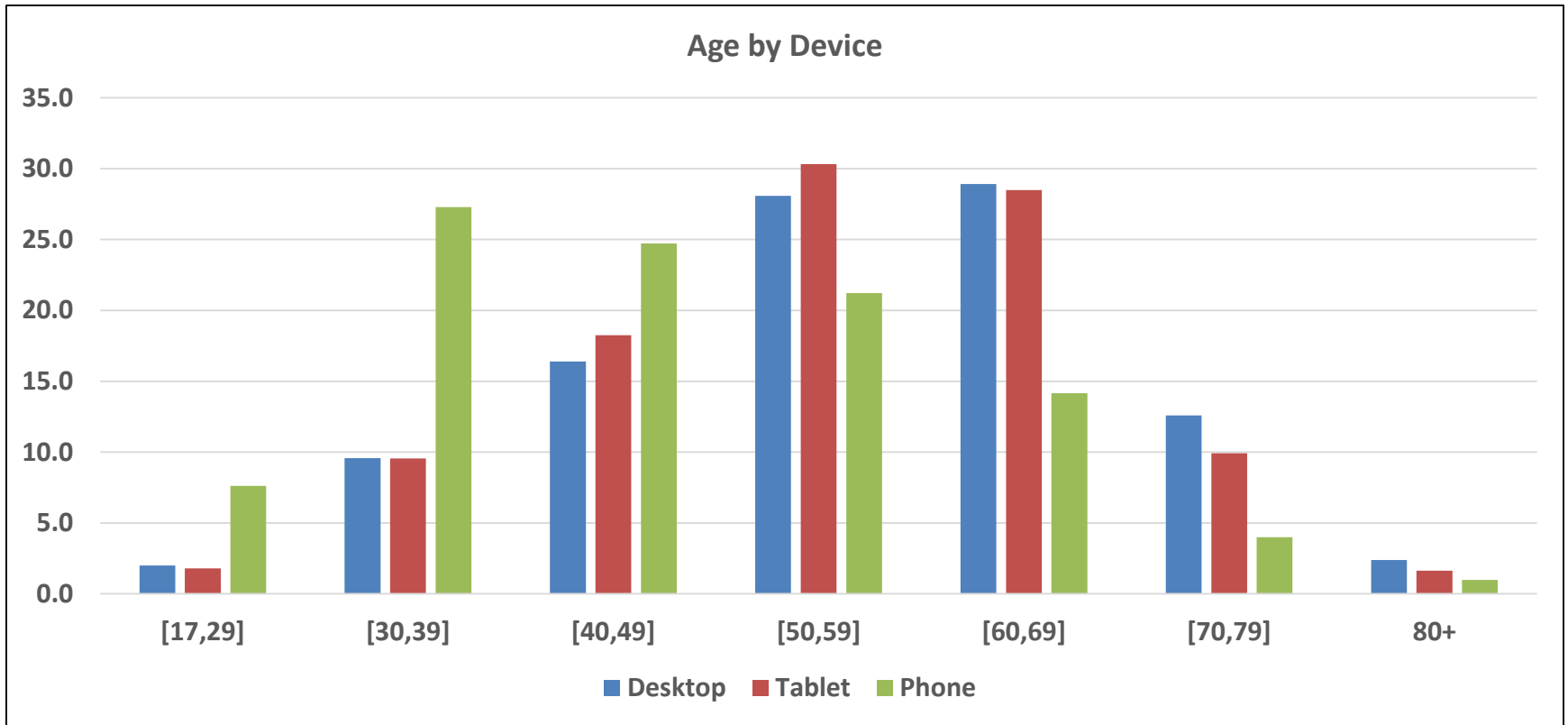
# Recommendations and Next Steps

- Continue to work on trying to understand what operator/operation characteristics are attributed to web reporting.

- Optimize web design for most common and current platforms.

- Almost 41% of respondents triggered a warning message.
  - Additional review and testing should be conducted to determine how best to collect difficult items in the web instrument.
  - Evaluate the effectiveness of warning messages which alert them to enter a response.

- Investigate through cognitive testing and/or usability testing questions where respondents frequently changed their answers.

# Contact Information

Robyn.Sirkis@nass.usda.gov


Pam.McGovern@nass.usda.gov