



The Role of Survey Paradata in a Federal Statistical Agency's Commitment to Quality

John Finamore
2019 FedCASIC Workshop
April 16, 2019

National Science Foundation
National Center for Science and Engineering Statistics
www.nsf.gov/statistics



Outline

- Role and responsibilities of federal statistical agencies
- Commitment to quality
- Considering paradata as part of this commitment
- Examples of paradata in NCSES's National Survey of College Graduates



*Role and Responsibilities of
Federal Statistical Agencies*



Definition of a federal statistical agency

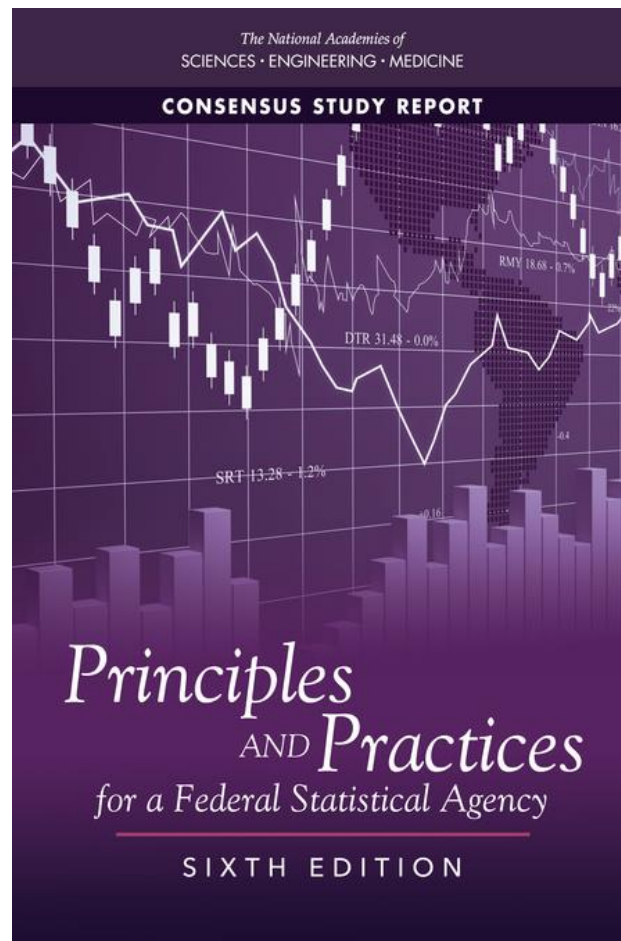
A federal statistical agency is a component of the federal government whose principal function is the compilation and analysis of data and the dissemination of information for statistical purposes.

(National Academies of Sciences, Engineering, and Medicine, 2017)

Documents that provide guidance for federal statistical agencies

- *Principles and Practices for a Federal Statistical Agency* (National Academies, 2017)
- *Statistical Policy Directive No. 1, Fundamental Responsibilities* (OMB, 2014)
- *Statistical Policy Directive No. 2, Standards and Guidelines* (OMB, 2006)
- *Fundamental Principles of Official Statistics* (United Nations Statistical Commission, 2014)
- *The European Statistics Code of Practice* (European Statistical System Committee, 2011)

Principles and Practices for a Federal Statistical Agency: Sixth Edition



Principles for a federal statistical agency

- Principle 1: Relevance to policy issues
- Principle 2: Credibility among data users
- Principle 3: Trust among data providers
- Principle 4: Independence from political and other undue external influence

Practices 1-7 for a federal statistical agency

- Practice 1: A clearly defined and well-accepted mission
- Practice 2: Necessary authority to protect independence
- Practice 3: Use of multiple data sources to meet needs
- Practice 4: Openness about data sources and limitations
- Practice 5: Wide dissemination of data
- Practice 6: Cooperation with data users
- Practice 7: Respect for the privacy of data providers



Practices 8-13 for a federal statistical agency

- Practice 8: Confidentiality protection for data providers
- Practice 9: Commitment to quality
- Practice 10: An active research program
- Practice 11: Professional advancement of staff
- Practice 12: A strong internal and external evaluation program
- Practice 13: Coordination and collaboration with other statistical agencies

Practices for a federal statistical agency

- Practice 1: A clearly defined and well-accepted mission
- Practice 2: Necessary authority to protect independence
- Practice 3: Use of multiple data sources to meet needs
- Practice 4: Openness about data sources and limitations
- Practice 5: Wide dissemination of data
- **Practice 9: Commitment to quality**
- Practice 6: Cooperation with data users
- Practice 7: Respect for the privacy of data providers
- Practice 8: Confidentiality protection for data providers
- Practice 9: Commitment to quality and professional standards of practice
- Practice 10: An active research program
- Practice 11: Professional advancement of staff
- Practice 12: A strong internal and external evaluation program
- Practice 13: Coordination and collaboration with other statistical agencies



Commitment to Quality



Activities demonstrating a commitment to quality

1. Discuss possible sources of error
2. Develop quality assurance efforts
3. Keep up to date on advances
4. Seek internal and external feedback



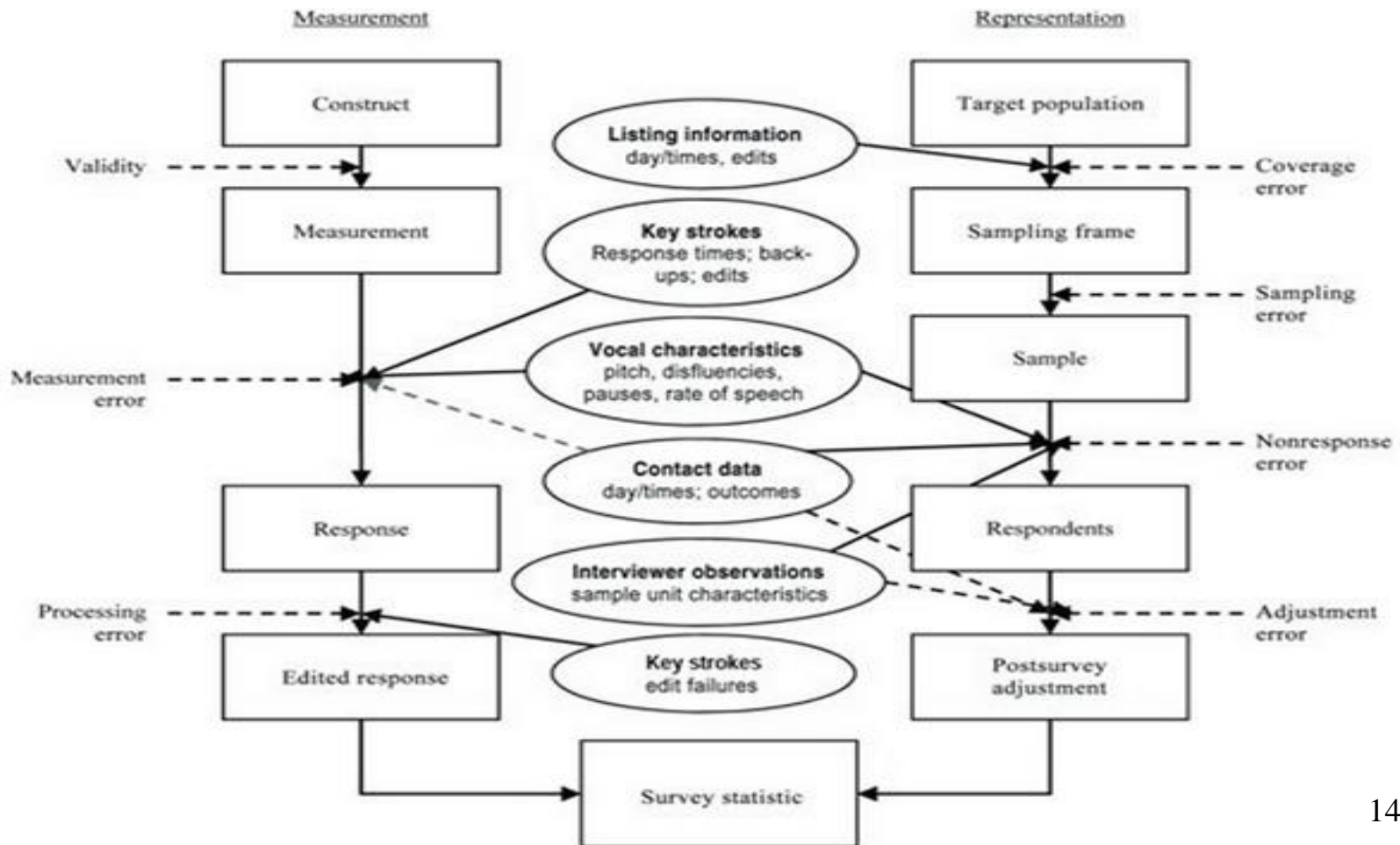
*Considering Paradata as Part of the
Commitment to Quality*

Definition of paradata

- Additional data that can be captured during the process of producing a survey estimate (Kreuter, 2013)
- Captured at all stages of the survey process
- Numerous possible uses
- Examples include interview observations, web survey navigation information, and contact history

Considering paradata in CQ activity 1

(Discuss possible sources of error)



Considering paradata in CQ activity 2

(Develop quality assurance efforts)

- Nonresponse error investigation
Paradata examples: contact history, interview observations
- Measurement error investigation
Paradata examples: time stamp, mouse clicks, vocal characteristics
- Contact/response rate improvement
Paradata examples: listing information, contact history, breakoffs
- Data collection intervention improvement
Paradata examples: contact history, interview observations



Considering paradata in CQ activity 3 *(Keep up to date on advances)*

- Web survey user experience
- Web survey measurement error reduction
- Data collection operational efficiency
- Adaptive design monitoring and intervention

Considering paradata in CQ activity 4 *(Seek internal and external feedback)*

- Interagency working groups
- Conferences and workshops
- Peer-reviewed journals





*Examples of Paradata in
NCSES's National Survey of College Graduates*



National Survey of College Graduates

- Biennial survey of U.S. residing college graduates
- Repeated cross-sectional survey
- Sequential mode offering (web first)
- 6 month data collection period
- 70% response rate in 2017
- 78% used web to respond in 2017

Ex #1: Paradata to inform response rate improvement

- Paradata: Device type (web survey paradata)
- Issue: Percent of web response by smartphones nearly tripled between cycles

Device type	Respondent device usage	
	2013	2015
Smartphone	1.7%	5.1%
Tablet	9.8%	8.3%
Computer/Other	88.5%	86.6%

- Issue: The survey completion time for smartphone far exceeded that of other web respondents

Device Type	2015 median completion time (minutes)
Smartphone	33.6
Tablet	29.9
Computer/Other	26.9

Ex #1: Paradata to inform response rate improvement

- Paradata: Device type (web survey paradata)
- Issue: Percent of web response by smartphones nearly tripled between cycles

Device type	Respondent device usage	
	2013	2015
Smartphone	1.7%	5.1%
Tablet	9.8%	8.3%
Computer/Other	88.5%	86.6%

- Issue: The survey completion time for smartphone far exceeded that of other web respondents

Device Type	2015 median completion time (minutes)
Smartphone	33.6
Tablet	29.9
Computer/Other	26.9

Ex #1: Paradata to inform response rate improvement

- Paradata: Device type (web survey paradata)
- Issue: Percent of web response by smartphones nearly tripled between cycles

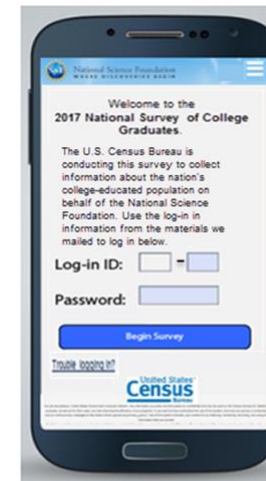
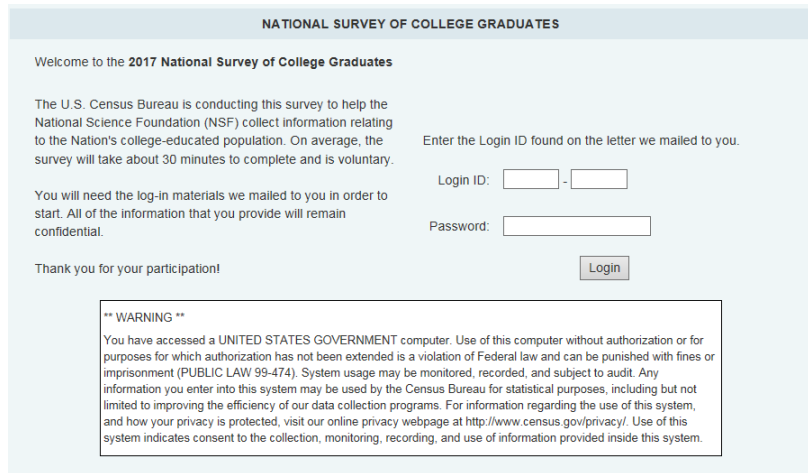
Device type	Respondent device usage	
	2013	2015
Smartphone	1.7%	5.1%
Tablet	9.8%	8.3%
Computer/Other	88.5%	86.6%

- Issue: The survey completion time for smartphone far exceeded that of other web respondents

Device Type	2015 median completion time (minutes)
Smartphone	33.6
Tablet	29.9
Computer/Other	26.9

Ex #1: Paradata to inform response rate improvement (continued)

- Resolution: Mobile optimized the web instrument

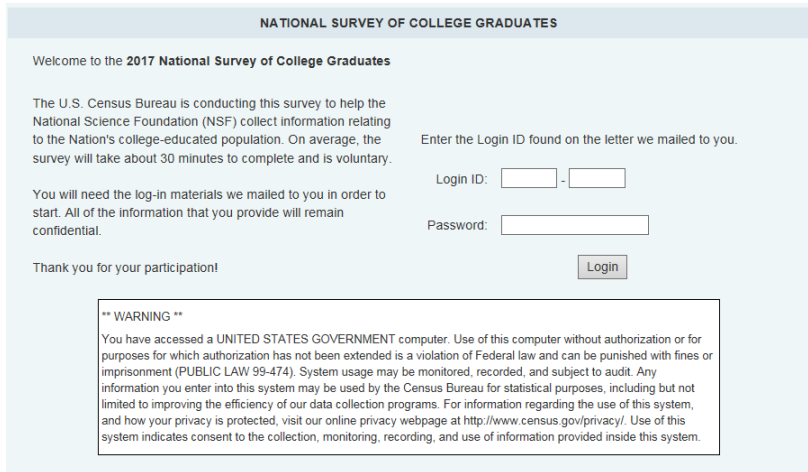


- Impact: Improved user experience

Device Type	Device usage		Median completion time (minutes)	
	2015	2017	2015	2017
Smartphone	5.1%	8.9%	33.6	24.5
Tablet	8.3%	6.4%	29.9	28.9
Computer/Other	86.6%	84.7%	26.9	24.9

Ex #1: Paradata to inform response rate improvement (continued)

- Resolution: Mobile optimized the web instrument

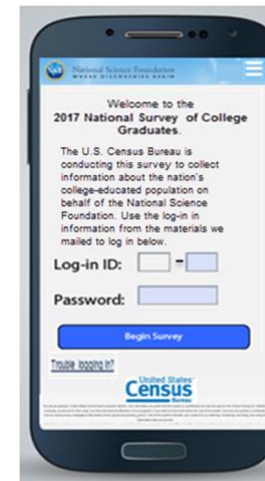
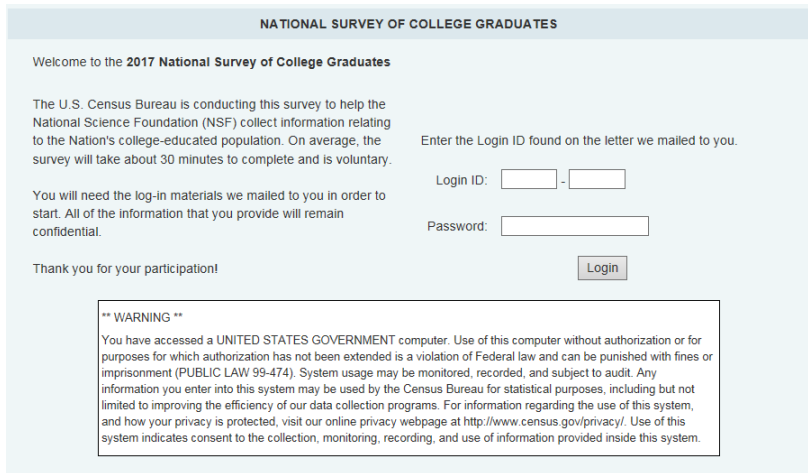


- Impact: Improved user experience

Device Type	Device usage		Median completion time (minutes)	
	2015	2017	2015	2017
Smartphone	5.1%	8.9%	33.6	24.5
Tablet	8.3%	6.4%	29.9	28.9
Computer/Other	86.6%	84.7%	26.9	24.9

Ex #1: Paradata to inform response rate improvement (continued)

- Resolution: Mobile optimized the web instrument

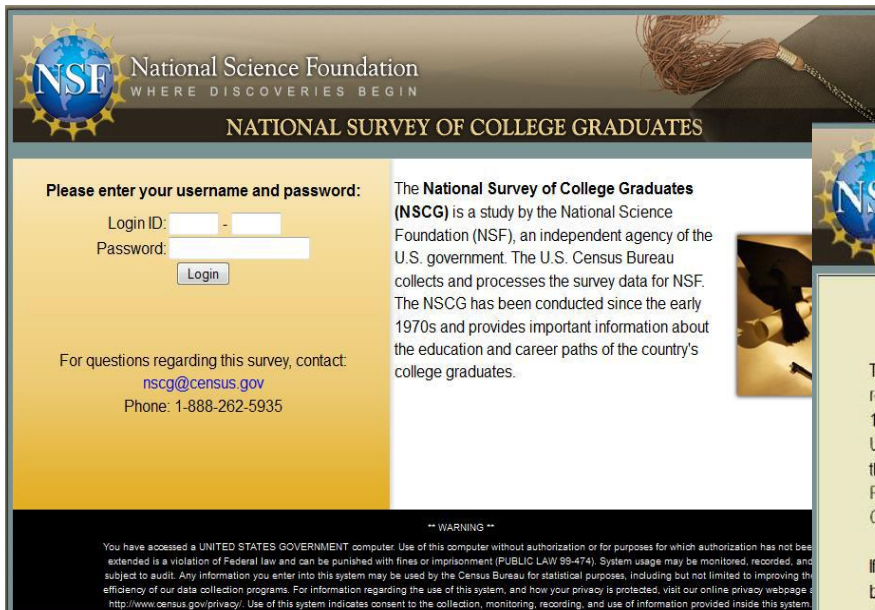


- Impact: Improved user experience

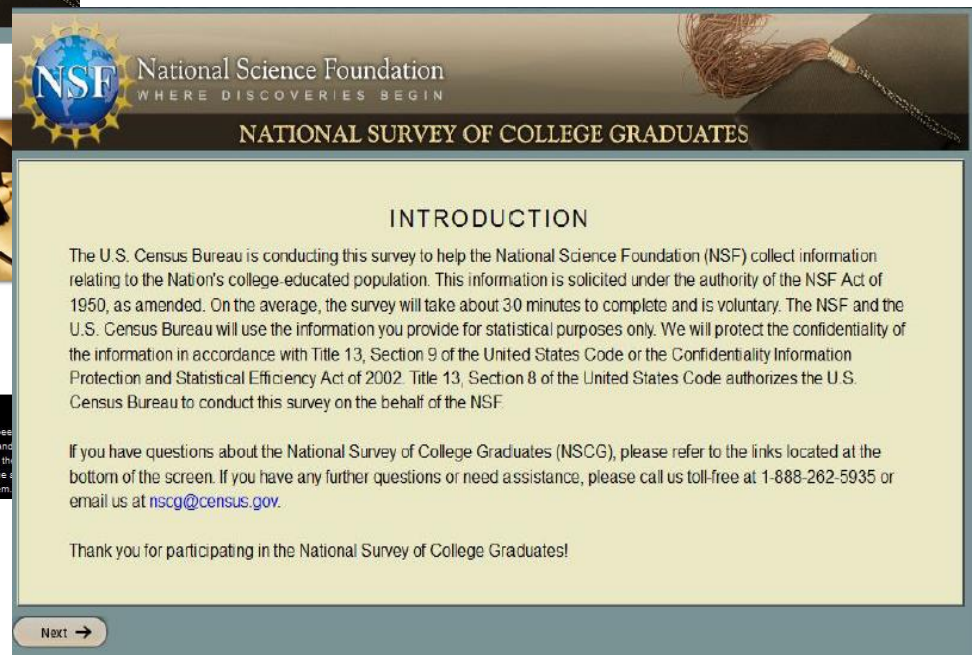
Device Type	Device usage		Median completion time (minutes)	
	2015	2017	2015	2017
Smartphone	5.1%	8.9%	33.6	24.5
Tablet	8.3%	6.4%	29.9	28.9
Computer/Other	86.6%	84.7%	26.9	24.9

Ex #2: Paradata to inform response rate improvement

- Paradata: Breakoff indicator (web survey paradata)
- Issue: High breakoff rate on introductory screen (13.7% of all breakoff)



The screenshot shows the login page for the National Survey of College Graduates. It features the NSF logo and the text "National Science Foundation WHERE DISCOVERIES BEGIN" and "NATIONAL SURVEY OF COLLEGE GRADUATES". The login section includes a "Please enter your username and password:" prompt, a "Login ID:" field, a "Password:" field, and a "Login" button. Below the login fields, there is contact information: "For questions regarding this survey, contact: nscg@census.gov Phone: 1-888-262-5935". To the right of the login fields, there is a paragraph of text: "The National Survey of College Graduates (NSCG) is a study by the National Science Foundation (NSF), an independent agency of the U.S. government. The U.S. Census Bureau collects and processes the survey data for NSF. The NSCG has been conducted since the early 1970s and provides important information about the education and career paths of the country's college graduates." At the bottom of the page, there is a "WARNING" message: "You have accessed a UNITED STATES GOVERNMENT computer. Use of this computer without authorization or for purposes for which authorization has not been extended is a violation of Federal law and can be punished with fines or imprisonment (PUBLIC LAW 98-474). System usage may be monitored, recorded, and subject to audit. Any information you enter into this system may be used by the Census Bureau for statistical purposes, including but not limited to improving the efficiency of our data collection programs. For information regarding the use of this system, and how your privacy is protected, visit our online privacy webpage at http://www.census.gov/privacy/. Use of this system indicates consent to the collection, monitoring, recording, and use of information provided inside this system."



The screenshot shows the introductory screen for the National Survey of College Graduates. It features the NSF logo and the text "National Science Foundation WHERE DISCOVERIES BEGIN" and "NATIONAL SURVEY OF COLLEGE GRADUATES". The main heading is "INTRODUCTION". The text reads: "The U.S. Census Bureau is conducting this survey to help the National Science Foundation (NSF) collect information relating to the Nation's college-educated population. This information is solicited under the authority of the NSF Act of 1950, as amended. On the average, the survey will take about 30 minutes to complete and is voluntary. The NSF and the U.S. Census Bureau will use the information you provide for statistical purposes only. We will protect the confidentiality of the information in accordance with Title 13, Section 9 of the United States Code or the Confidentiality Information Protection and Statistical Efficiency Act of 2002. Title 13, Section 8 of the United States Code authorizes the U.S. Census Bureau to conduct this survey on the behalf of the NSF." Below the introduction, there is a paragraph: "If you have questions about the National Survey of College Graduates (NSCG), please refer to the links located at the bottom of the screen. If you have any further questions or need assistance, please call us toll-free at 1-888-262-5935 or email us at nscg@census.gov." At the bottom of the page, there is a "Thank you for participating in the National Survey of College Graduates!" message and a "Next" button with a right arrow.

Ex #2: Paradata to inform response rate improvement (continued)

- Resolution: Combined information from Introductory screen to log-in screen (removed non-actionable screen)

NATIONAL SURVEY OF COLLEGE GRADUATES

Welcome to the **2017 National Survey of College Graduates**

The U.S. Census Bureau is conducting this survey to help the National Science Foundation (NSF) collect information relating to the Nation's college-educated population. On average, the survey will take about 30 minutes to complete and is voluntary.

You will need the log-in materials we mailed to you in order to start. All of the information that you provide will remain confidential.

Thank you for your participation!

Enter the Login ID found on the letter we mailed to you.

Login ID: -

Password:

**** WARNING ****

You have accessed a UNITED STATES GOVERNMENT computer. Use of this computer without authorization or for purposes for which authorization has not been extended is a violation of Federal law and can be punished with fines or imprisonment (PUBLIC LAW 99-474). System usage may be monitored, recorded, and subject to audit. Any information you enter into this system may be used by the Census Bureau for statistical purposes, including but not limited to improving the efficiency of our data collection programs. For information regarding the use of this system, and how your privacy is protected, visit our online privacy webpage at <http://www.census.gov/privacy/>. Use of this system indicates consent to the collection, monitoring, recording, and use of information provided inside this system.

- Impact: Total breakoffs reduced (7.7% (2015) → 0.5% (2017))
- Impact: Percentage of breakoffs on first screen reduced (13.7% (2015) → 1.5% (2017))

Ex #3: Paradata to inform measurement error improvement

- Paradata: Previous clicks (web survey paradata)
- Issue: Items with an “other, specify” response option had a high percentage of previous clicks

A22. Did any of the following factors influence your decision to work in an area outside the field of your highest degree?

Please refer to the work in the principal job you held during the week of February 1, 2015.

Select Yes or No for each item.

	Yes	No
1. Pay, promotion opportunities	<input type="radio"/>	<input type="radio"/>
2. Working conditions (e.g., hours, equipment, working environment)	<input type="radio"/>	<input type="radio"/>
3. Job location	<input type="radio"/>	<input type="radio"/>
4. Change in career or professional interests	<input type="radio"/>	<input type="radio"/>
5. Family-related reasons (e.g., children, spouse's job moved)	<input type="radio"/>	<input type="radio"/>
6. Job in highest degree field not available	<input type="radio"/>	<input type="radio"/>
7. Some other factor	<input type="radio"/>	<input type="radio"/>

← Previous

Next →

Ex #3: Paradata to inform measurement error improvement

- Paradata: Previous clicks (web survey paradata)
- Issue: Items with an “other, specify” response option had a high percentage of previous clicks

A22. Did any of the following factors influence your decision to work in an area outside the field of your highest degree?

Please refer to the work in the principal job you held during the week of February 1, 2015.

Select Yes or No for each item.

	Yes	No
1. Pay, promotion opportunities	<input type="radio"/>	<input type="radio"/>
2. Working conditions (e.g., hours, equipment, working environment)	<input type="radio"/>	<input type="radio"/>
3. Job location	<input type="radio"/>	<input type="radio"/>
4. Change in career or professional interests	<input type="radio"/>	<input type="radio"/>
5. Family-related reasons (e.g., children, spouse's job moved)	<input type="radio"/>	<input type="radio"/>
6. Job in highest degree field not available	<input type="radio"/>	<input type="radio"/>
7. Some other factor	<input type="radio"/>	<input type="radio"/>

← Previous Next →

Ex #3: Paradata to inform measurement error improvement

- Paradata: Previous clicks (web survey paradata)
- Issue: Items with an “other, specify” response option had a high percentage of previous clicks

A22. Did any of the following factors influence your decision to work in an area outside the field of your highest degree?
Please refer to the work in the principal job you held during the week of February 1, 2015.
Select Yes or No for each item.

	Yes	No
1. Pay, promotion opportunities	<input type="radio"/>	<input type="radio"/>
2. Working conditions	<input type="radio"/>	<input type="radio"/>
3. Job location	<input type="radio"/>	<input type="radio"/>
4. Change in career interests	<input type="radio"/>	<input type="radio"/>
5. Family-related issues	<input type="radio"/>	<input type="radio"/>
6. Job in highest degree field	<input type="radio"/>	<input type="radio"/>
7. Some other factor	<input type="radio"/>	<input type="radio"/>

Part A - Principal Employer

A22a. Please specify the reason that influenced your decision to work in an area outside the field of your highest degree.
Please refer to the work in the principal job you held during the week of February 1, 2015.

← Previous Next →

Ex #3: Paradata to inform measurement error improvement (continued)

- Resolution: Removed “other, specify” follow-up screen and added textbox to original item

Part A - Principal Job

A22. Did any of the following factors influence your decision to work in an area outside the field of your highest degree?
Please refer to the work in the principal job you held during the week of February 1, 2017.

Select Yes or No for each item.

	Yes	No
Pay, promotion opportunities	<input type="radio"/>	<input type="radio"/>
Working conditions (e.g., hours, equipment, working environment)	<input type="radio"/>	<input type="radio"/>
Job location	<input type="radio"/>	<input type="radio"/>
Change in career or professional interests	<input type="radio"/>	<input type="radio"/>
Family-related reasons (e.g., children, spouse's job moved)	<input type="radio"/>	<input type="radio"/>
Job in highest degree field not available	<input type="radio"/>	<input type="radio"/>
Some other factor, <i>Specify</i>	<input type="radio"/>	<input type="radio"/>

Previous Next

- Impact: Previous clicks reduced by 7 percentage points
- Impact: Item-level completion time reduced by 30 seconds

Ex #3: Paradata to inform measurement error improvement (continued)

- Resolution: Removed “other, specify” follow-up screen and added textbox to original item

Part A - Principal Job

A22. Did any of the following factors influence your decision to work in an area outside the field of your highest degree?
Please refer to the work in the principal job you held during the week of February 1, 2017.

Select Yes or No for each item.

	Yes	No
Pay, promotion opportunities	<input type="radio"/>	<input type="radio"/>
Working conditions (e.g., hours, equipment, working environment)	<input type="radio"/>	<input type="radio"/>
Job location	<input type="radio"/>	<input type="radio"/>
Change in career or professional interests	<input type="radio"/>	<input type="radio"/>
Family-related reasons (e.g., children, spouse's job moved)	<input type="radio"/>	<input type="radio"/>
Job in highest degree field not available	<input type="radio"/>	<input type="radio"/>
Some other factor, <i>Specify</i>	<input type="radio"/>	<input type="radio"/>

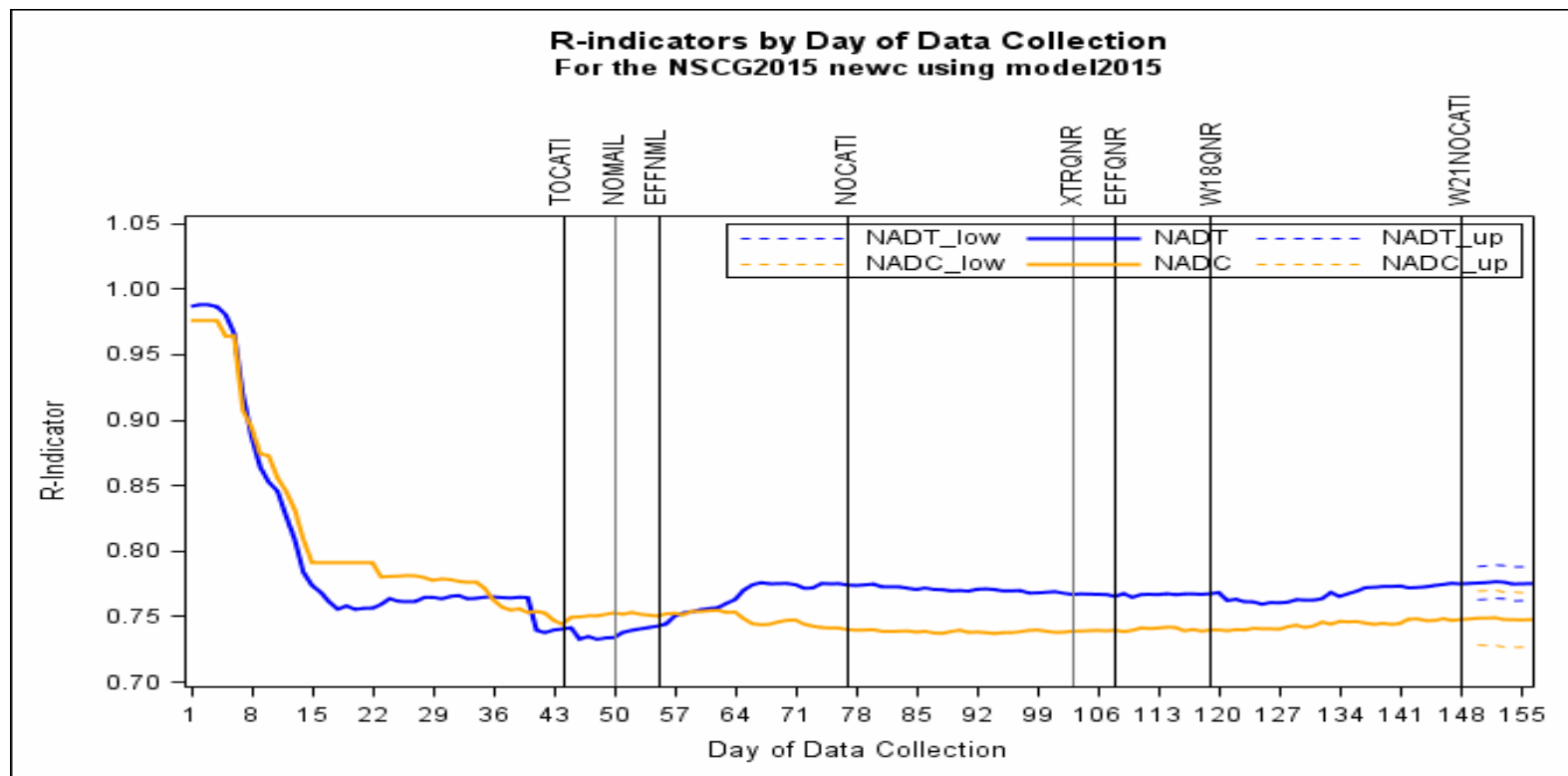
- Impact: Previous clicks reduced by 7 percentage points
- Impact: Item-level completion time reduced by 30 seconds

Ex #4: Paradata to inform nonresponse error improvement

- Paradata: Contact history
- Issue: Underrepresentation of key population groups (minorities, women, individuals with disabilities)
- Resolution: Adaptive design interventions used contact history to model propensity to respond
- Impact: Increased representation of these groups as measured by the R-indicator

Ex #4: Paradata to inform nonresponse error improvement (continued)

- Impact: In the 2015 NSCG, when we started making adaptive design interventions, the *R-indicator point estimate* for the treatment groups became and stayed more representative throughout data collection



Concluding thoughts

- Paradata is a valuable tool to pinpoint areas for improvement and improve data quality
- We have evidence that these steps reduce respondent burden (and frustration)
- We have indirect evidence that these steps reduce the potential for measurement error and nonresponse error
- Further research is needed to show a direct link between these steps and an error reduction in our survey estimates



Thank you!

John Finamore
Program Director, Human Resources Statistics Program
National Center for Science and Engineering Statistics
National Science Foundation

jfinamor@nsf.gov
(703) 292-2258