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

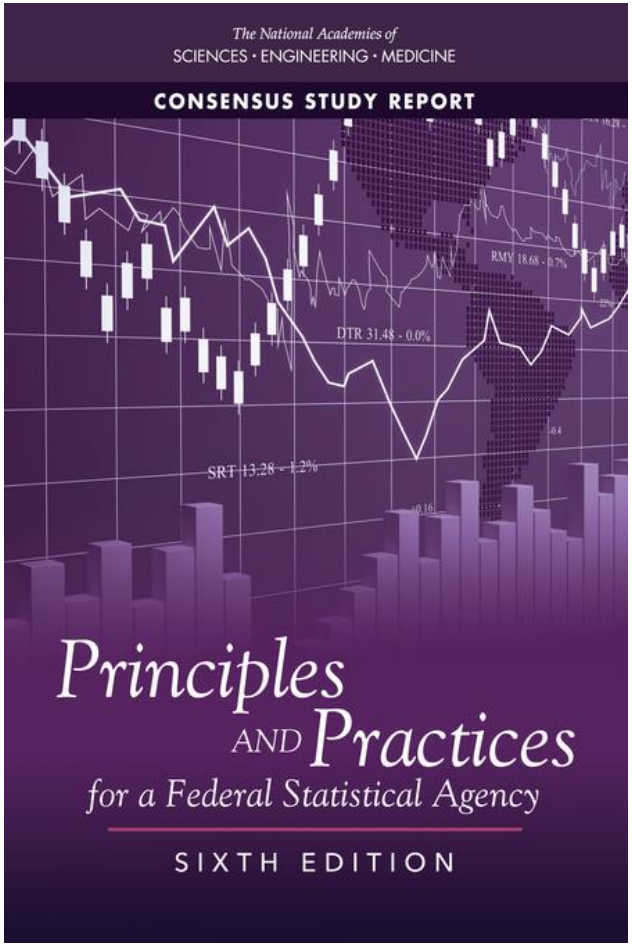
A federal statistical agency is a component of the federal government whose principal function is the compilation and analysis of data and the dissemination of information for statistical purposes.

(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 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

Practice 9: Commitment to quality

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

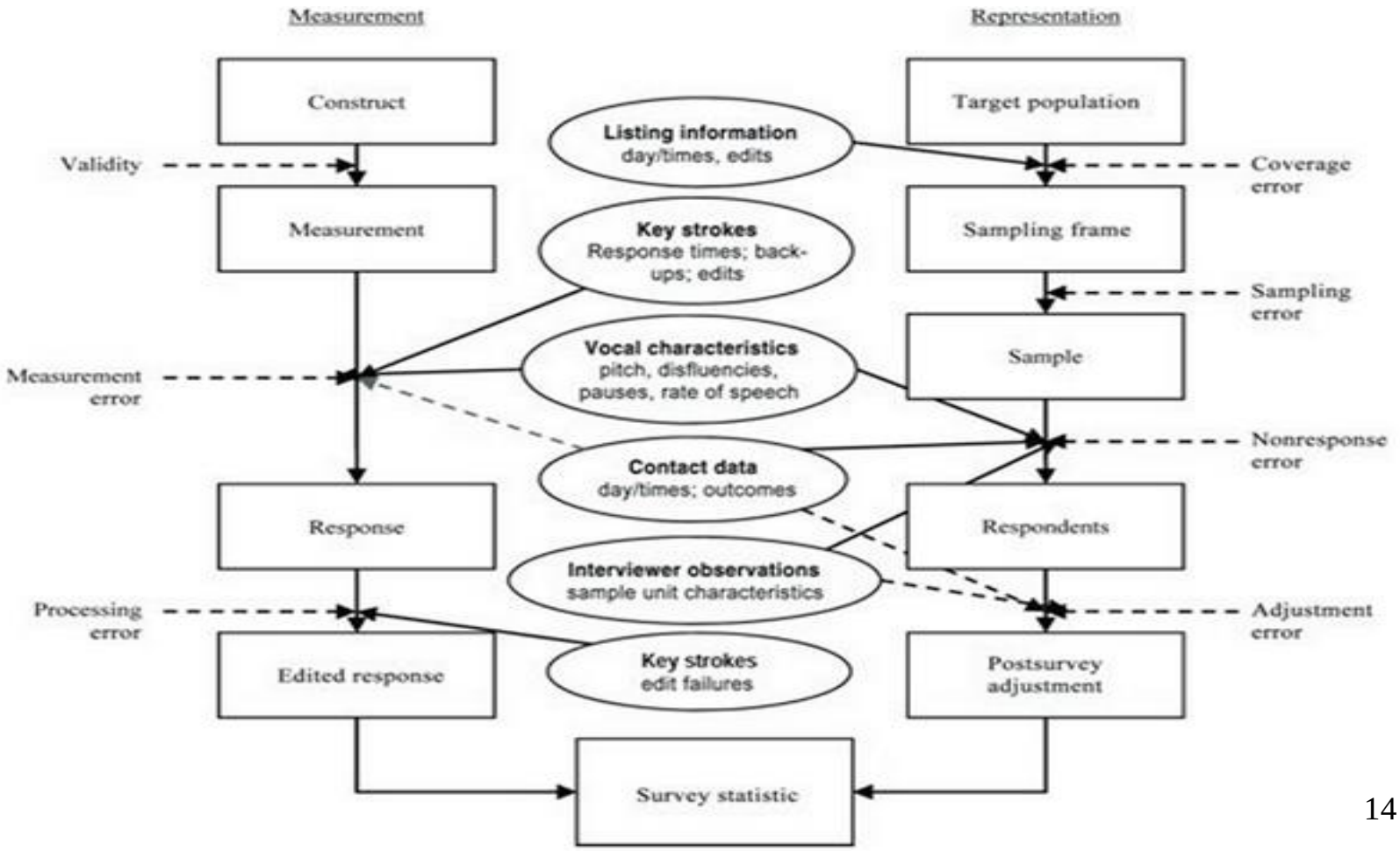
Considering Paradata as Part of the Commitment to Quality

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

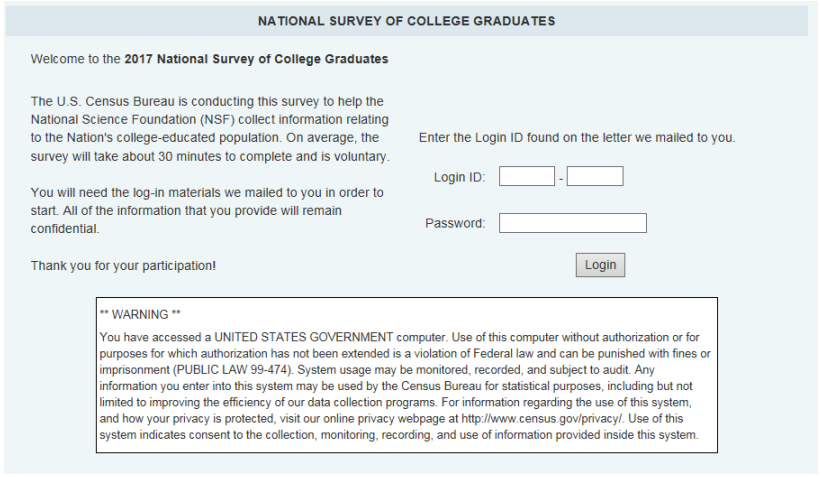
Device type	Respondent device usage 2013	2015
Smartphone	1.7%	5.1%
Tablet	9.8%	8.3%
Computer/Other	88.5%	86.6%

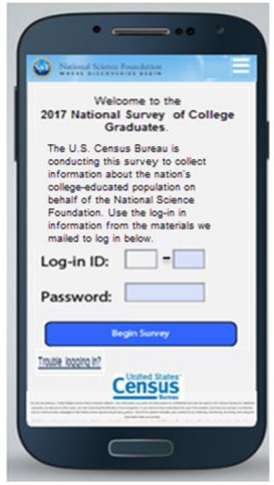
- Issue: The survey completion time for smartphone far exceeded that of other web respondents

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



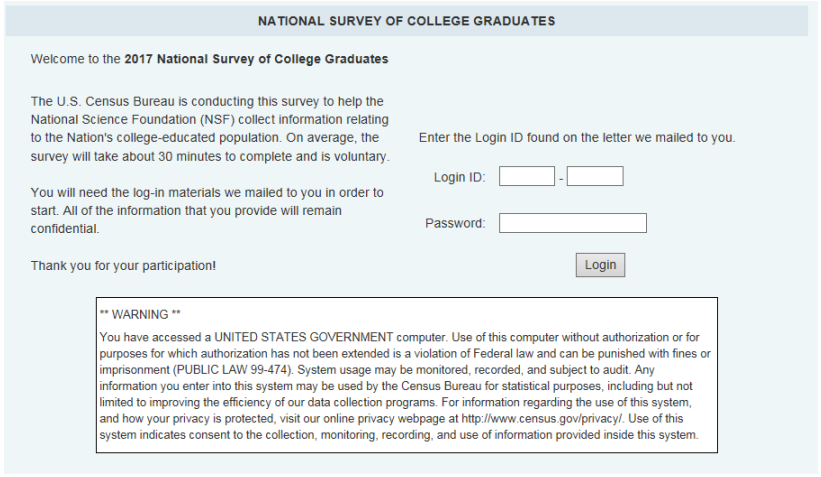


- Impact: Improved user experience

Device Type	Device usage 2015	Device usage 2017	Median completion time (minutes) 2015	Median completion time (minutes) 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

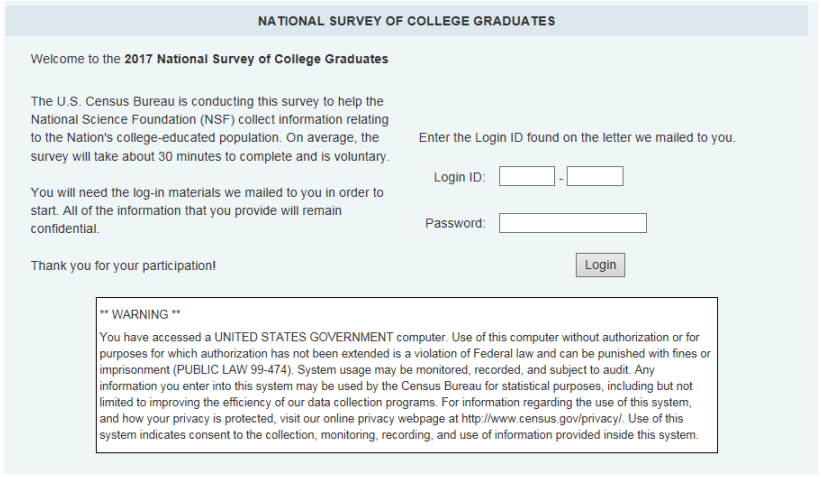


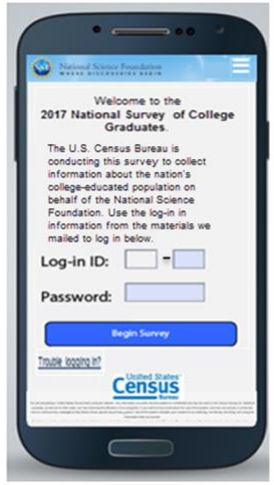
- Impact: Improved user experience

Device Type	Device usage 2015	Device usage 2017	Median completion time (minutes) 2015	Median completion time (minutes) 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



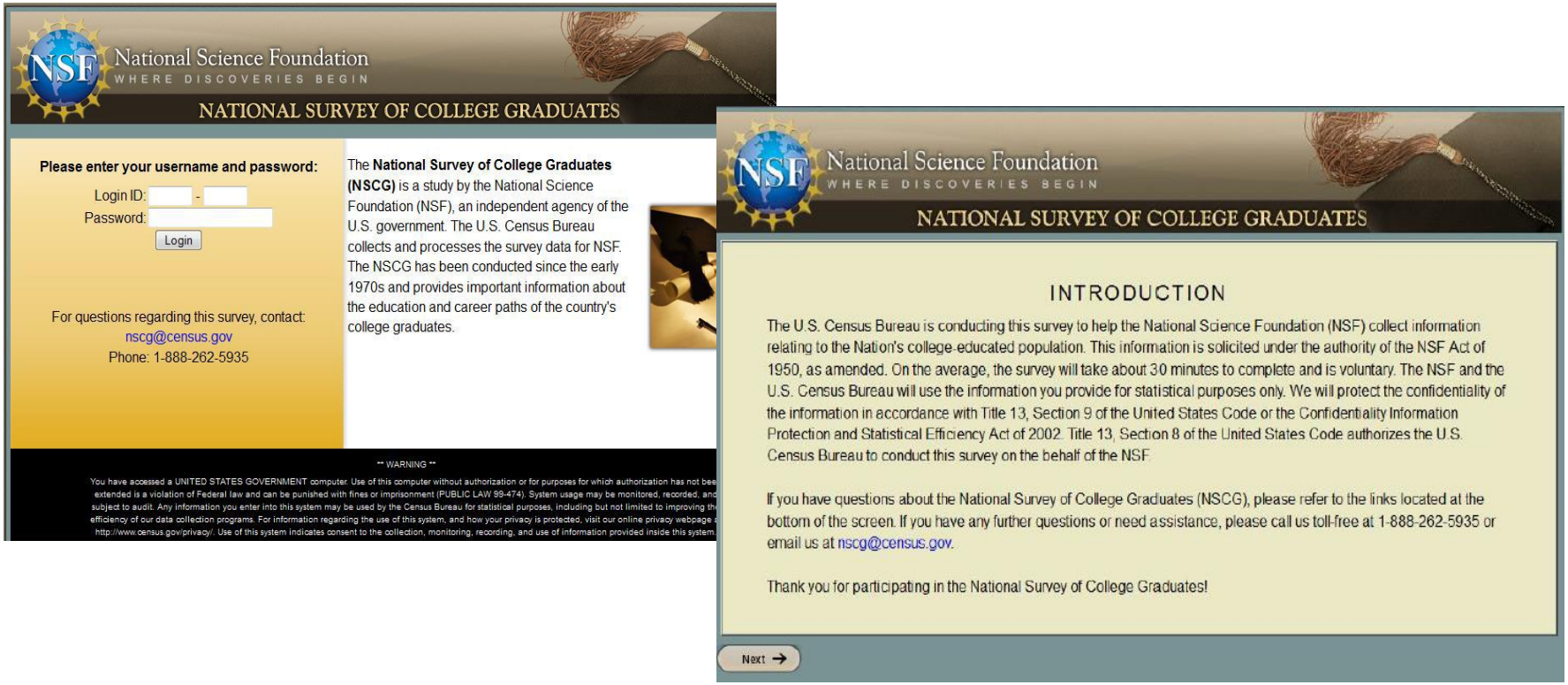


- Impact: Improved user experience

Device Type	Device usage 2015	Device usage 2017	Median completion time (minutes) 2015	Median completion time (minutes) 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

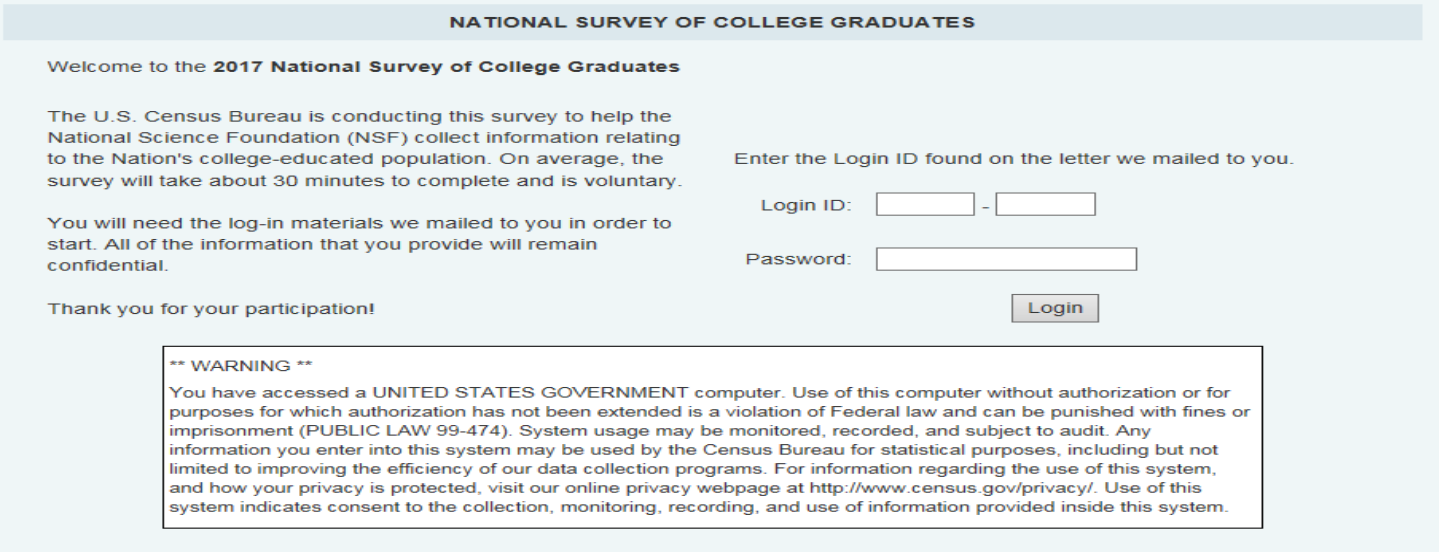
Ex #2: Paradata to inform response rate improvement

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



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

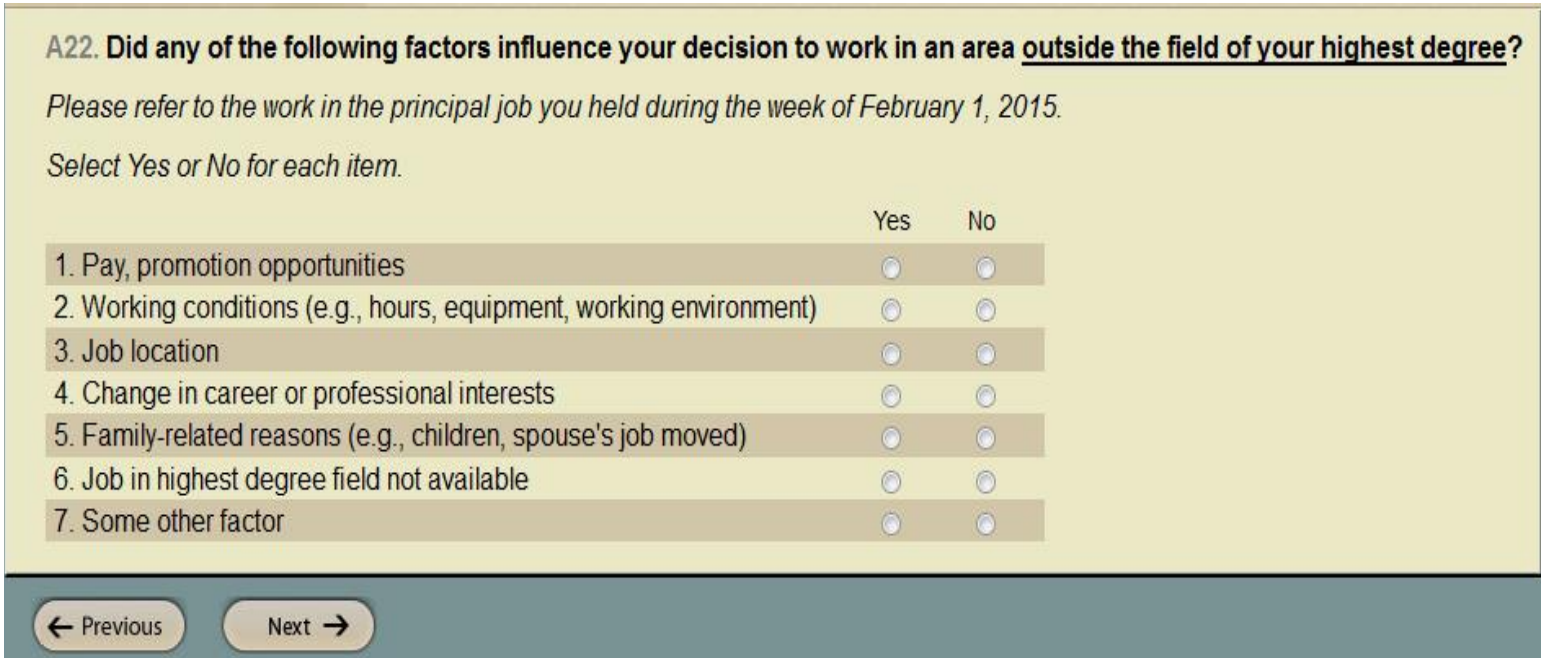
- Resolution: Combined information from Introductory screen to log-in screen (removed non-actionable screen)



- 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.

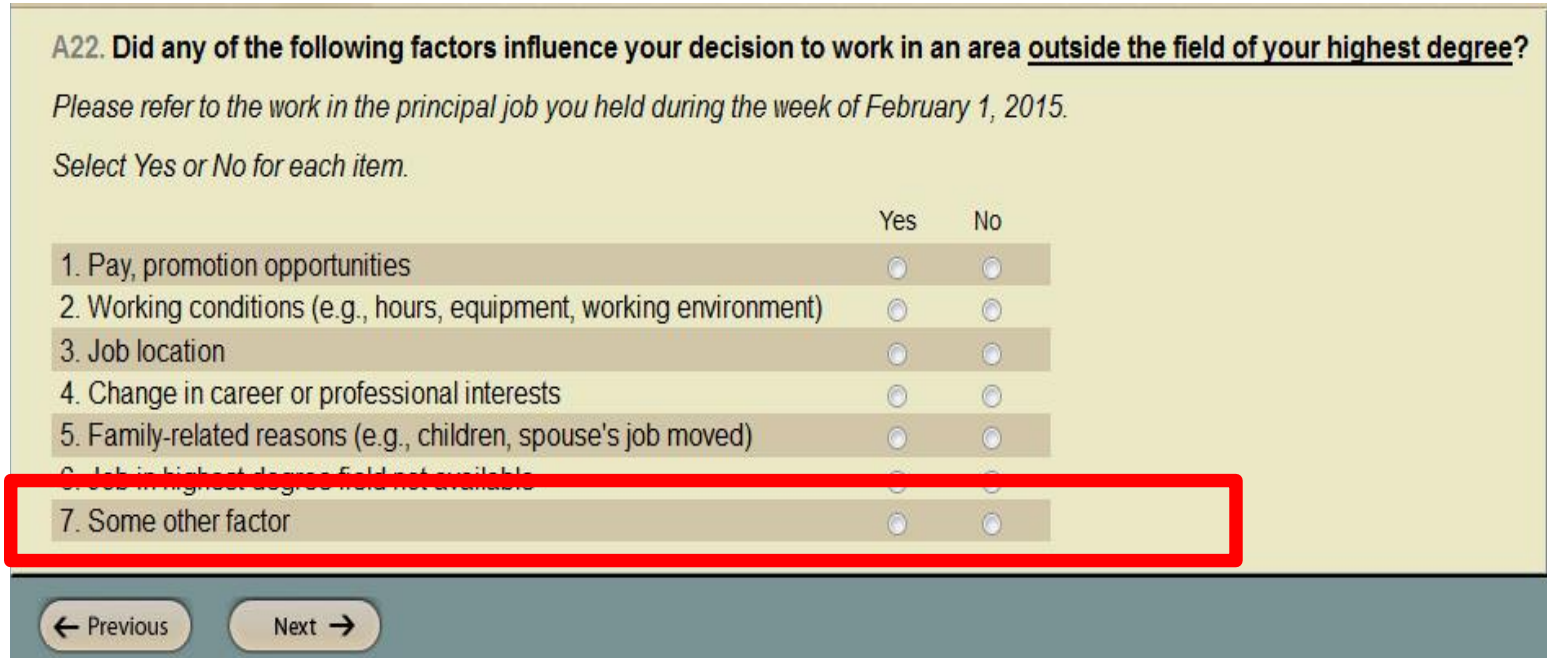
Select Yes or No for each item.

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

← Previous Next →

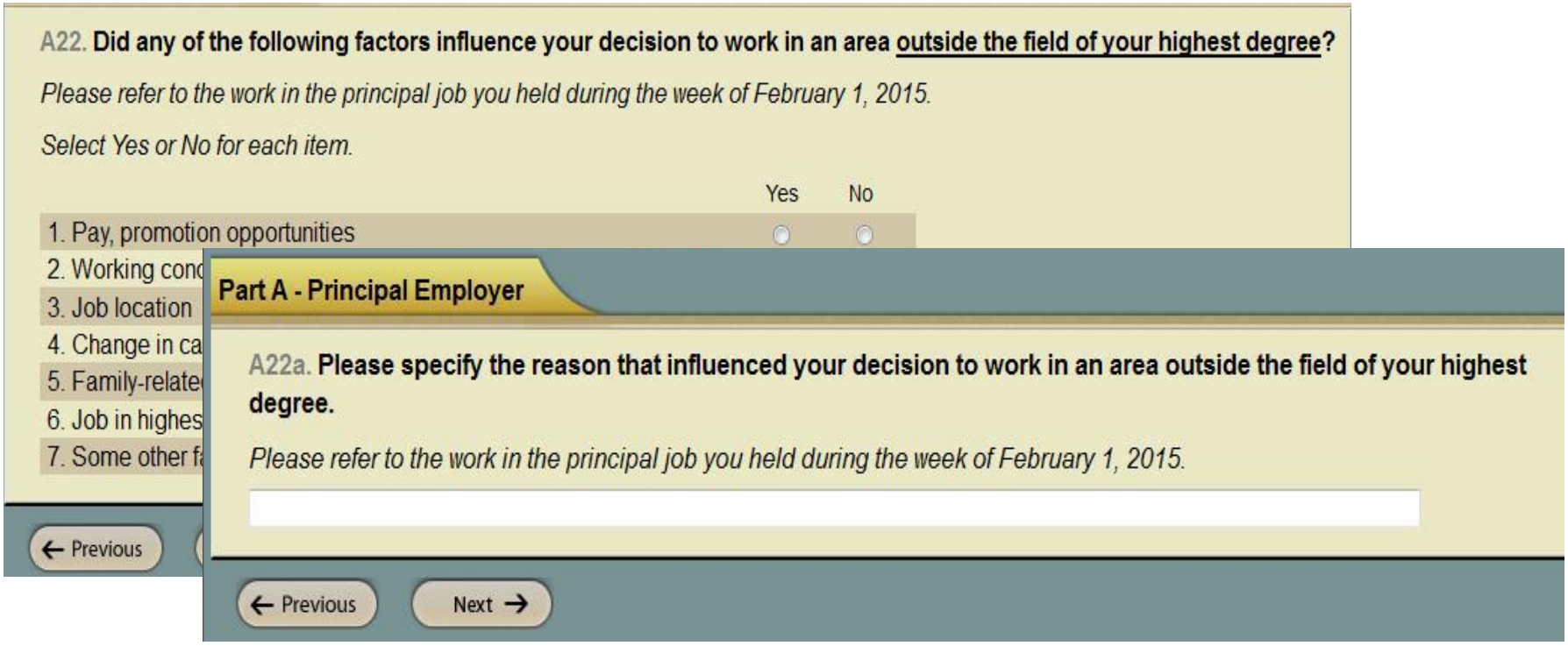
Ex #3: Paradata to inform measurement error improvement

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



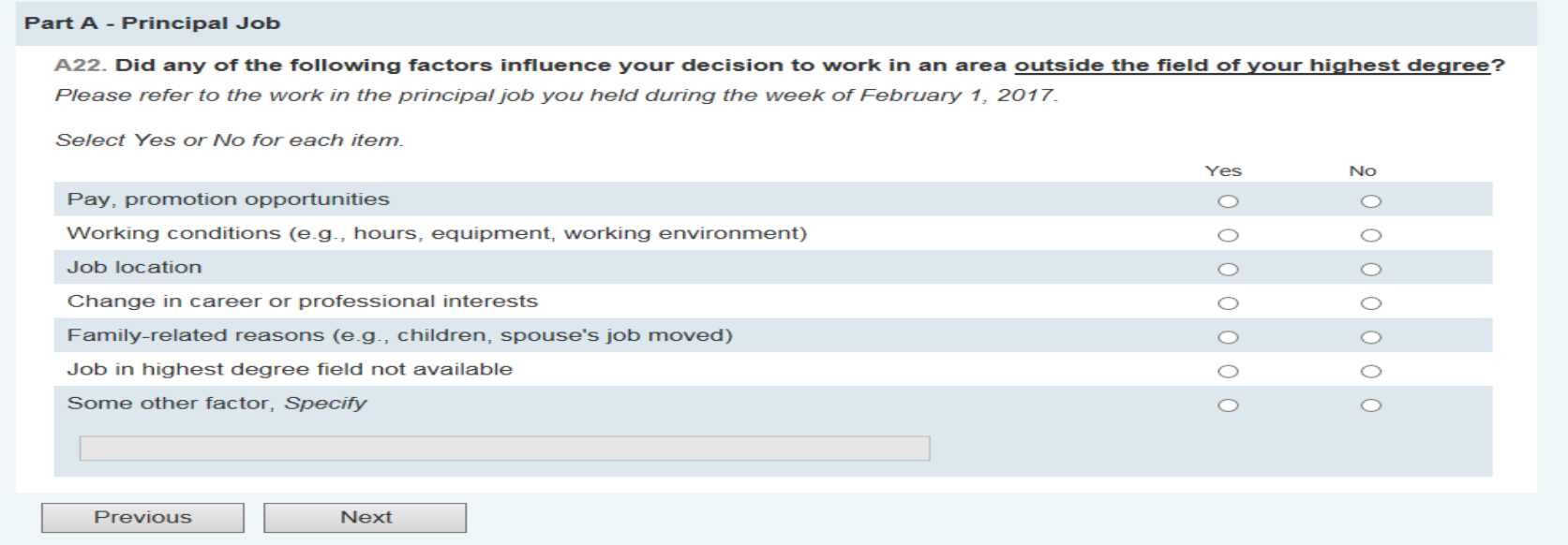
Ex #3: Paradata to inform measurement error improvement

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



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.

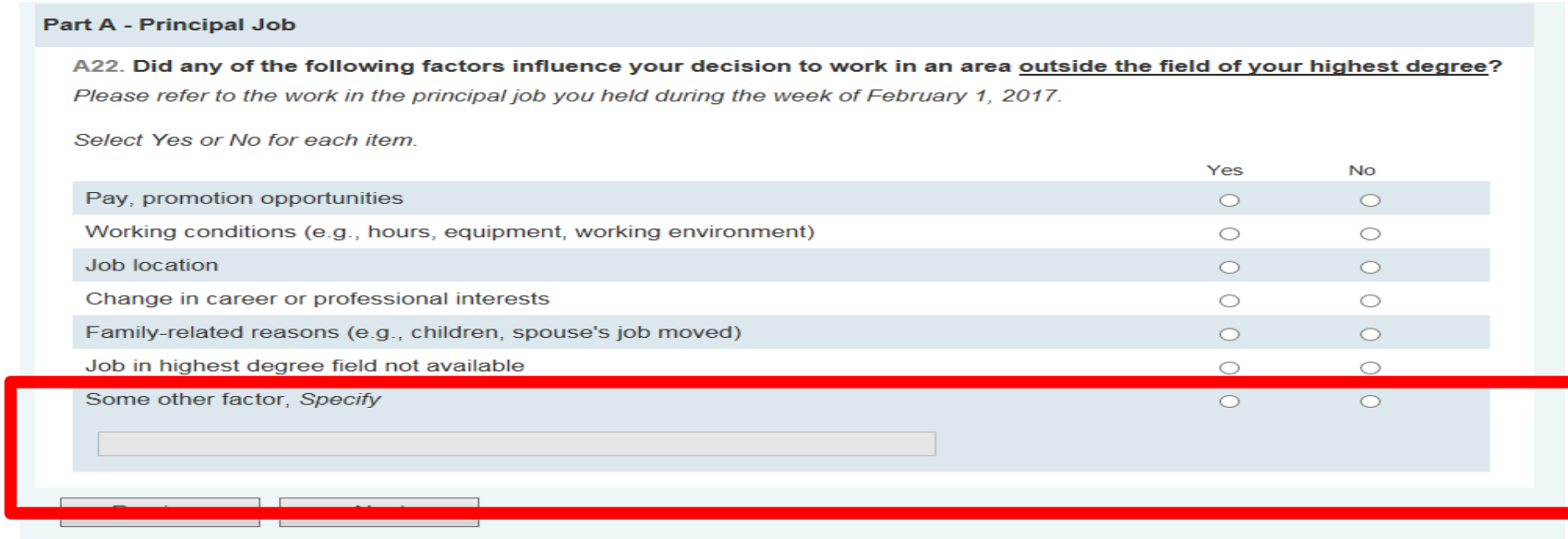
	Yes	No
Pay, promotion opportunities	○	○
Working conditions (e.g., hours, equipment, working environment)	○	○
Job location	○	○
Change in career or professional interests	○	○
Family-related reasons (e.g., children, spouse's job moved)	○	○
Job in highest degree field not available	○	○
Some other factor,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

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	○	○
Working conditions (e.g., hours, equipment, working environment)	○	○
Job location	○	○
Change in career or professional interests	○	○
Family-related reasons (e.g., children, spouse's job moved)	○	○
Job in highest degree field not available	○	○
Some other factor, *Specify*	○	○

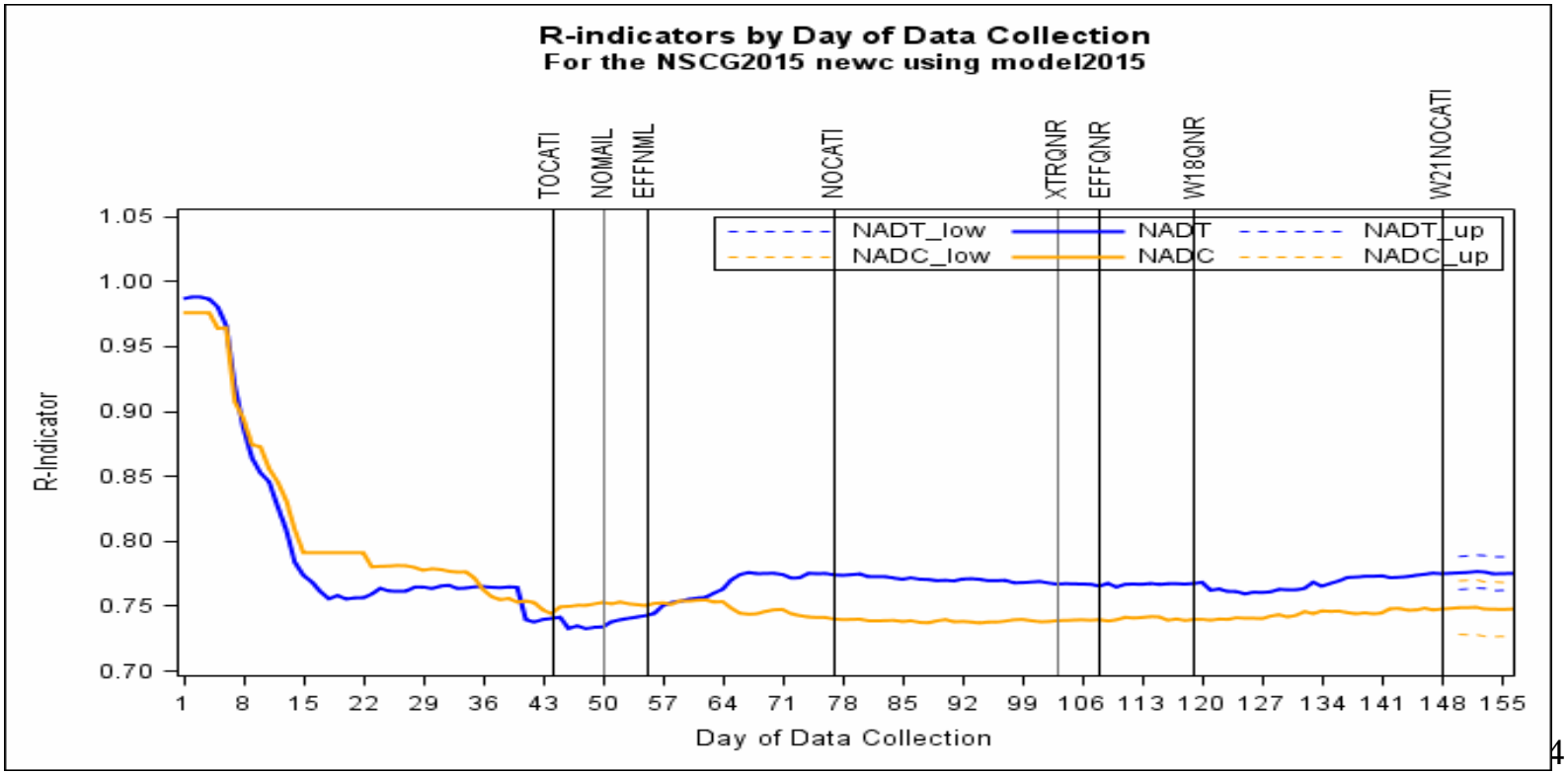
- Impact: Previous clicks reduced by 7 percentage points
- Impact: Item-level completion time reduced by 30 seconds

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