



Evaluating Data Collection Mode Options in the National Survey of College Graduates

*2012 Federal Committee on Statistical Methodology
Research Conference*

John Finamore

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



Disclaimer

This presentation is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

Any views expressed on statistical, methodological, technical, or operational issues are those of the author and not necessarily of the National Science Foundation.

All results included in this presentation are preliminary and have not been tested for statistical significance.



Acknowledgements

Thanks to Don Dillman for his encouragement to pursue this research and the insight he has provided on the direction of the analysis.

Thanks to the Census Bureau staff for their assistance in the use of the NSCG data files and the derivation of NSCG data collection costs.



Overview

- Definitions
- Background
- Research Questions
- Methodology
- Results
- Conclusions and Discussion



Definitions



Definitions

- Mode – The data collection method by which a respondent completes a survey questionnaire
- This presentation discusses three data collection modes
 - **Web** refers to the method of completing a survey online
 - **Mail** refers to the method of completing a paper survey questionnaire and returning it by mail
 - **Telephone** refers to the method of a telephone interviewer reading the survey questions to the respondent



Background



A Survey Goal – Increase Web Response

- A desired goal for surveys is to capture responses over the web as an alternative to using mail or telephone
- A web response mode offers the potential of capturing a demographically different set of respondents
- A web response mode also offers potential cost savings over mail and telephone through elimination of:
 - Telephone interviewer staffing and long distance costs
 - Mail questionnaire printing and data capture costs



Challenges of Increasing Web Response

- Incorporating a web response mode can have an adverse effect on the overall response rate
 - Offering a concurrent choice (web or mail) has not improved overall response or response by web
 - Other options (e.g., web-only, web-plus-mail) have increased response by web, but have not achieved the overall response of a mail-only design
- Difficulty in encouraging response by web when using postal contacts rather than email contacts
- The computer literacy of the survey population affects the willingness to respond by web



How to Incorporate a Web Response Mode?

- What is an optimal design for incorporating a web response mode?
- Limited research exists on using web, mail, and telephone in a mixed-mode design
- This presentation discusses a survey that provided an opportunity to investigate ways to incorporate these three data collection modes
- This research examines results from four data collection designs that used these modes in various combinations



Research Questions



Research Questions

1. Does one mode, when used individually, result in a response rate increase over the other modes?
2. When sequentially offering additional modes, does the response rate vary based on the initial mode?
3. When sequentially offering additional modes, does the proportion of response by mode vary based on the initial mode?



Research Questions (continued)

4. Are there differences in the demographic composition of the responding sample across data collection designs?
5. Are there differences in the data collection costs across data collection designs?



Methodology



National Survey of College Graduates

- A biennial survey conducted by the U.S. Census Bureau for the survey sponsor, the National Science Foundation
- NSCG target population is the college-educated U.S. resident population under the age of 76
- The NSCG derived its sample of 65,195 cases from the 2009 American Community Survey
- This sample represents the population of 56 million college graduates



Mode Experiment

- Historically, the NSCG has used a mixed-mode data collection design (1990s – mail, telephone, personal visit; 2000s – mail, telephone)
- The 2010 cycle marked the introduction of the web data collection mode to the NSCG
- To determine the most effective manner to use the web mode along with mail and telephone, the NSCG included a mode experiment
- “Effective” is measured by overall response, response by mode, demographic representativeness, and cost



Mode Experiment Treatment Groups

- Eligibility criteria include a valid address and valid phone number
- Of the 65,195 cases, 60,875 were eligible and were randomly assigned to the four treatment groups
 - Web first group (n = 5,000)
 - Mail first group (n = 5,000)
 - Telephone first group (n = 3,500)
 - Choice group (n = 47,375)
- Each treatment group used a different data collection design



Data Collection

- The data collection effort only used mail-based contacts and telephone-based contacts (i.e., no email contacts)
- The contact strategy was similar across the treatment groups, but the available modes differed
- The NSCG data collection effort included five distinct phases:
 - Invitation Phase (weeks 1-4)
 - Initial Reminder Phase (weeks 5-7)
 - Secondary Reminder Phase (weeks 8-14)
 - Production CATI Phase (weeks 15-31)
 - Incentive Phase (weeks 32-42)



Contact Methodology

<u>Treatment Group</u>	<u>Invitation Phase</u>				
	Week 0	Week 1	Week 2	Week 3	Week 4
Web First	Prenotice letter	Web invite letter	Postcard		
Mail First	Prenotice letter	Mail invite letter & questionnaire	Postcard		
Telephone First	Prenotice letter	Telephone calls			
Choice	Prenotice letter	Web/mail invite letter & questionnaire	Postcard		



Contact Methodology (continued)

<u>Treatment Group</u>	<u>Initial Reminder Phase</u>		
	Week 5	Week 6	Week 7
Web First	Web invite letter	Automated phone call	
Mail First	Mail invite letter & questionnaire	Automated phone call	
Telephone First	Telephone invite letter	Telephone calls	
Choice			



Contact Methodology (continued)

<u>Treatment Group</u>	<u>Secondary Reminder Phase</u>				
	Week 8	Week 9	Weeks 10-12	Week 13	Week 14
Web First	Web/mail invite letter & questionnaire	Postcard		Automated phone call	
Mail First	Web/mail invite letter & questionnaire	Postcard		Automated phone call	
Telephone First	Web/mail invite letter & questionnaire	Postcard	Telephone calls		
Choice	Web/mail invite letter & questionnaire	Postcard		Automated phone call	



Response Mode Options

	<u>Data Collection Phase</u>				
	Invitation Phase (Weeks 1-4)	Initial Reminder Phase (Weeks 5-7)	Secondary Reminder Phase (Weeks 8-14)	Production CATI Phase (Weeks 15-31)	Incentive Phase (Weeks 32-42)
Web First	Web	Web	Web	Web	Web
			Mail	Mail	
Mail First	Mail	Mail	Web	Web	Web
			Mail	Mail	
Telephone First	Telephone	Telephone	Web	Web	Web
			Mail	Mail	
Choice	Web	Web	Web	Web	Web
	Mail	Mail	Mail	Mail	



Results

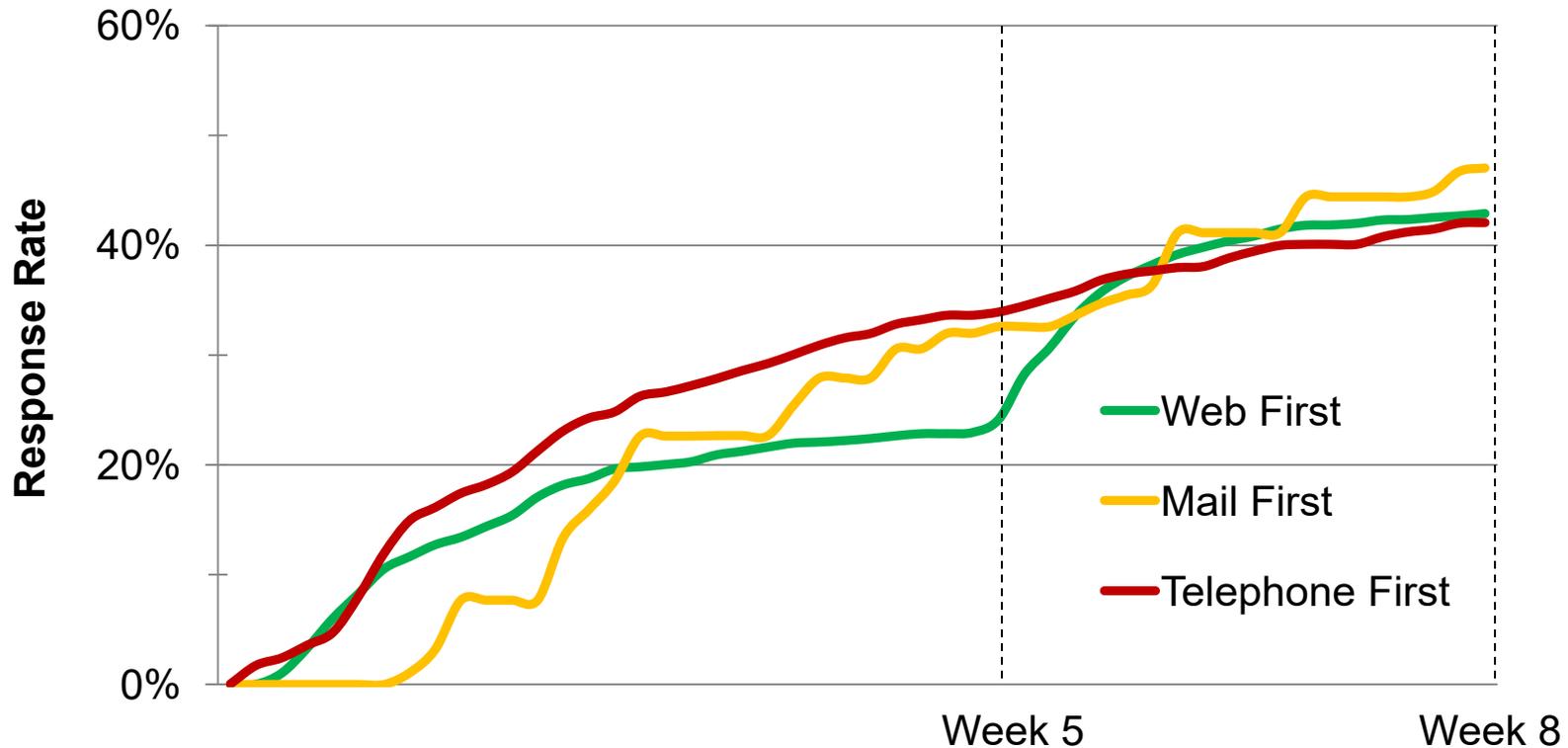
Disclaimer:

All results included in this presentation are preliminary and have not been tested for statistical significance



Results

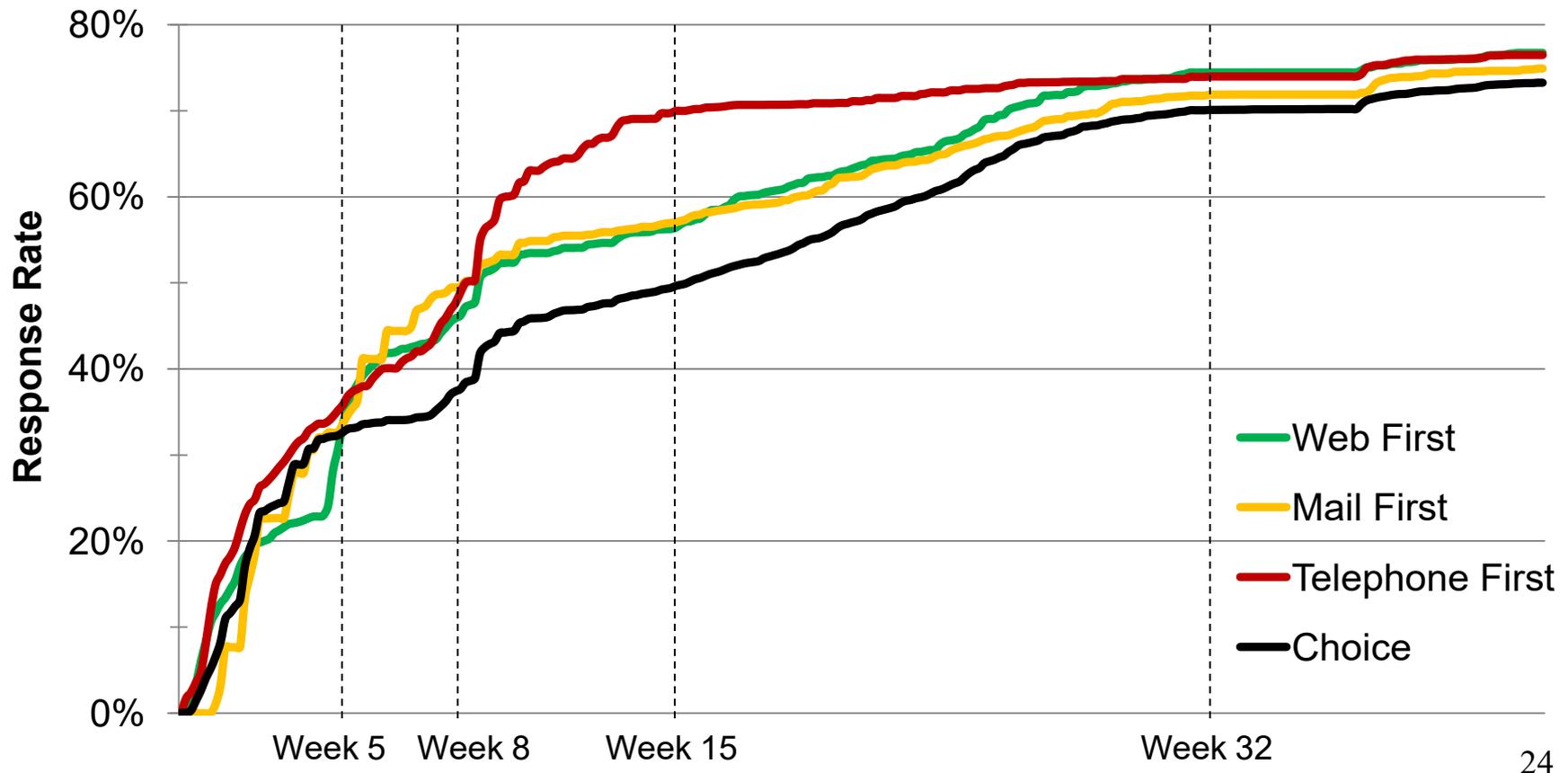
1. Does one mode, when used individually, result in a response rate increase over the other modes?





Results (continued)

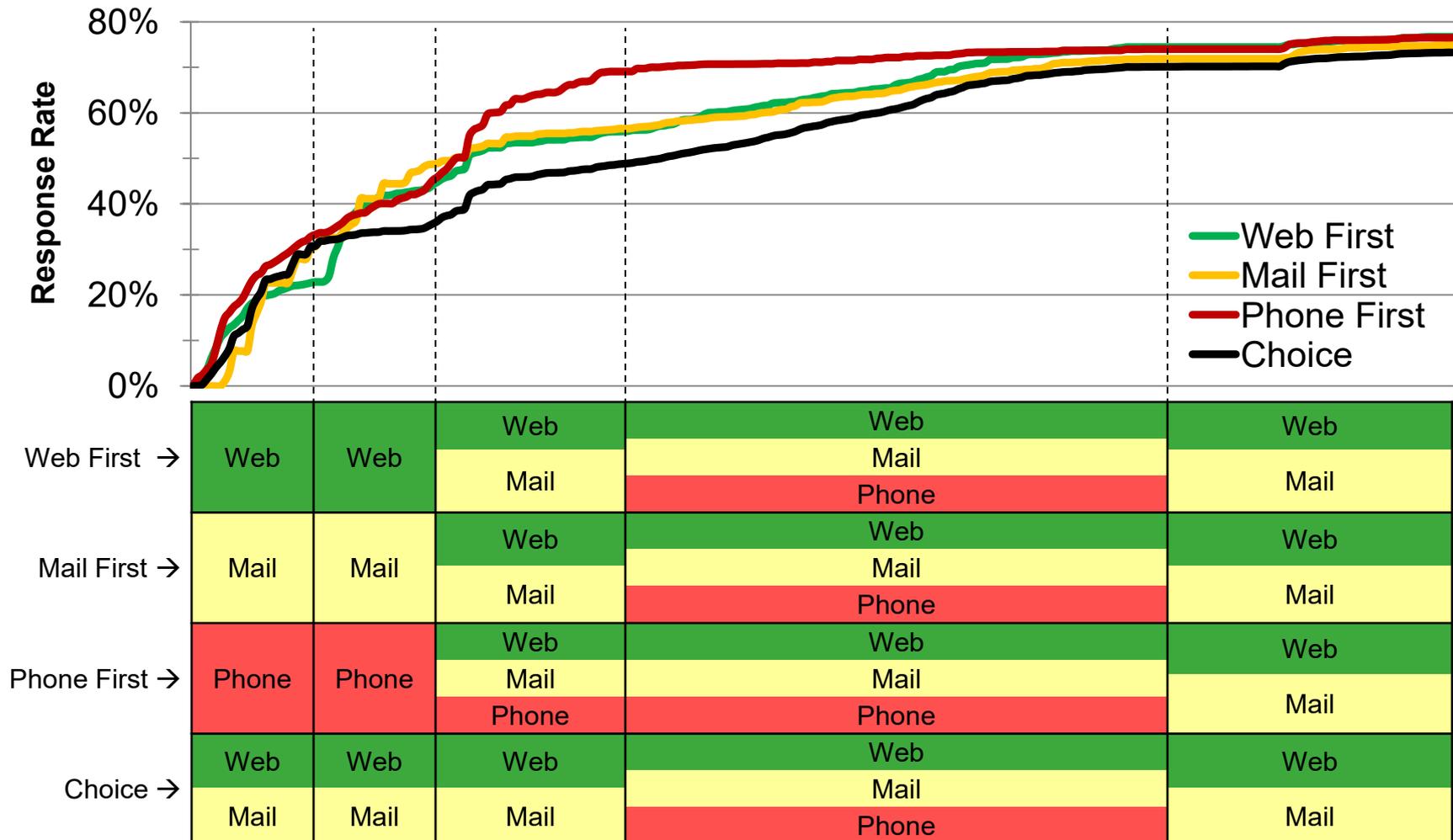
2. When sequentially offering additional modes, does the response rate vary based on the initial mode?





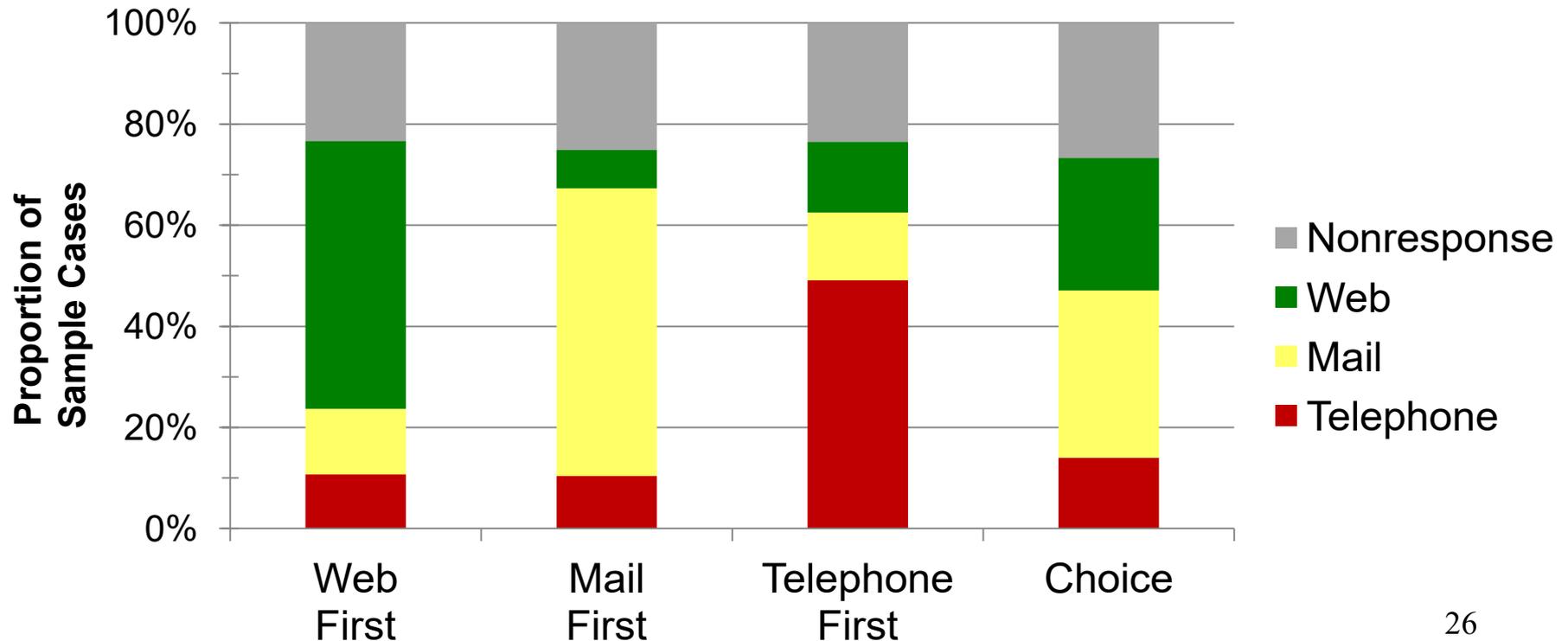
Results (continued)

2. When sequentially offering additional modes, does the response rate vary based on the initial mode?



Results (continued)

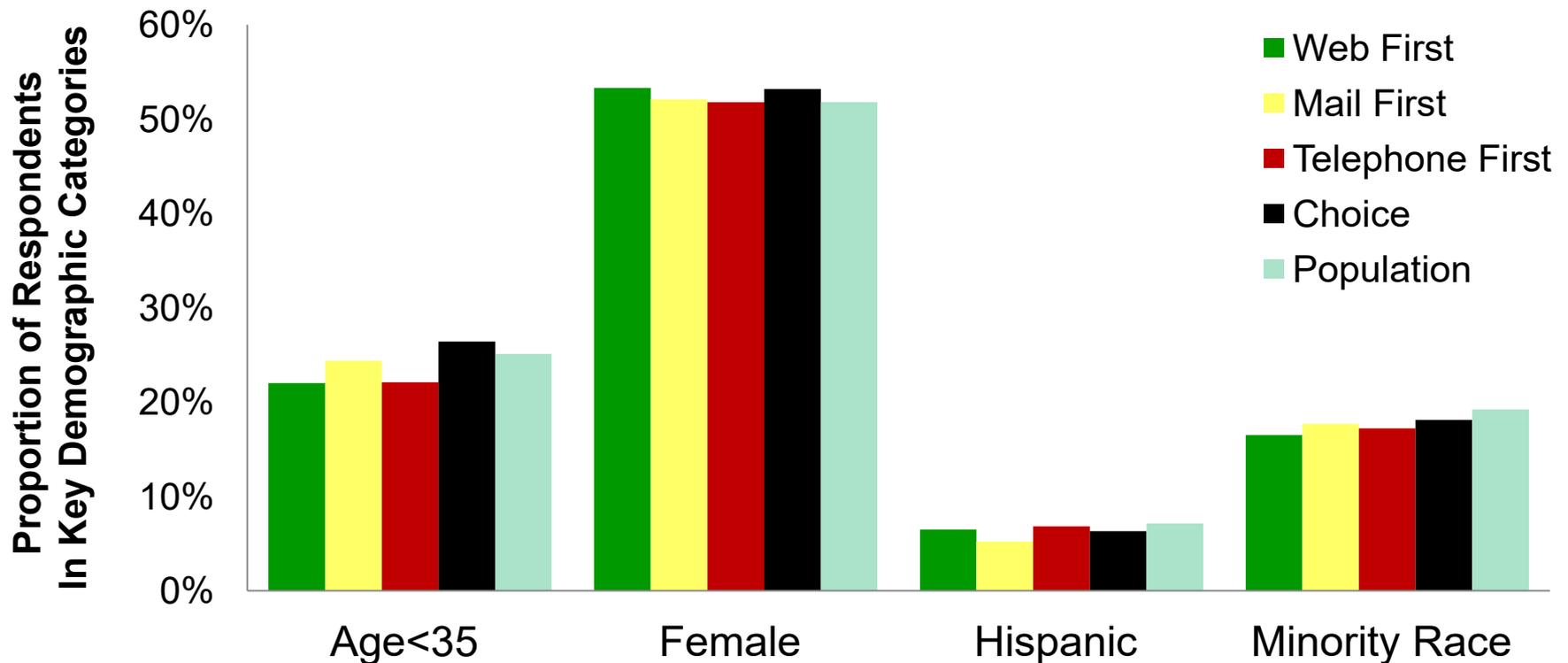
3. When sequentially offering additional modes, does the proportion of response by mode vary based on the initial mode?





Results (continued)

4. Are there differences in the demographic composition of the responding sample across data collection designs?





Results (continued)

5. Are there differences in the data collection costs across data collection designs?

Data Collection Costs at Week 8

Treatment Group	Data Collection Activities					Cost per Response	
	Instrument Development			Mail-based Contacts	Telephone Staff & Calls		Data Capture
	Web	Mail	Phone				
Web First	✓			✓		\$13.80	
Mail First		✓		✓		\$37.44	
Telephone First			✓	✓	✓	\$70.42	
Choice	✓	✓		✓		\$26.02	



Results (continued)

5. Are there differences in the data collection costs across data collection designs?

Data Collection Costs at the End of Data Collection

Treatment Group	Data Collection Activities						Cost per Response
	Instrument Development			Mail-based Contacts	Telephone Staff & Calls	Data Capture	
	Web	Mail	Phone				
Web First	✓	✓	✓	✓	✓	✓	\$48.28
Mail First	✓	✓	✓	✓	✓	✓	\$66.04
Telephone First	✓	✓	✓	✓	✓	✓	\$74.77
Choice	✓	✓	✓	✓	✓	✓	\$58.38



Conclusions

- When applied to a college-educated population, the web first data collection design:
 - Increases (or maintains) overall response
 - Increases response by web
 - Captures a demographically representative sample
 - Has a lower cost per response



Discussion

- Based on these results, general population surveys should consider including experiments within their own surveys to identify population subgroups that could benefit from a web first data collection strategy
- If the survey is able to identify such a group, the survey could consider a hybrid data collection design
 - Web first design for the identified group (or geographic areas with a high percent of identified group)
 - Mail first design for all other cases



Contact Information

John Finamore

National Center for Science and Engineering Statistics

National Science Foundation

jfinamor@nsf.gov

(703) 292-2258