

# National Household Education Surveys Program of 2019

## Methodological Experiments Report

# National Household Education Surveys Program of 2019

## Methodological Experiments Report

**June 2022**

**Rebecca Medway**

**Mahi Megra**

**Michael Jackson**

**Zoe Padgett**

**Danielle Battle**

American Institutes for Research

**Michelle McNamara**

*Project Officer*

National Center for Education Statistics

**U.S. Department of Education**

Miguel A. Cardona  
*Secretary*

**Institute of Education Sciences**

Mark Schneider  
*Director*

**National Center for Education Statistics**

Peggy Carr  
*Commissioner*

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

NCES, IES, U.S. Department of Education  
Potomac Center Plaza  
550 12th Street, SW  
Washington, DC 20202

**June 2022**

The NCES Home Page address is <https://nces.ed.gov>.

The NCES Publications and Products address is <https://nces.ed.gov/pubsearch>.

This report was prepared for the National Center for Education Statistics under Contract No. ED-IES-12-D-0002 with the American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

**Suggested Citation**

Medway, R., Megra, M., Jackson, M., Padgett, Z. and Battle, D. (2022). *National Household Education Surveys Program of 2019: Methodological Experiments Report* (NCES 2022-001). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2022001>.

**Content Contact**

Michelle McNamara  
(202) 245-7033  
[Michelle.McNamara@ed.gov](mailto:Michelle.McNamara@ed.gov)

# Executive Summary

This report presents the results of several methodological experiments that were embedded in the 2019 administration of the National Household Education Surveys Program (NHES:2019). These experiments were organized around three key themes: (1) better understanding how the offered response modes affect response rates; (2) refining and updating web-push<sup>1</sup> screener-phase contact methods<sup>2</sup> based on findings from the initial web-push experimentation conducted in NHES:2016; and (3) attracting response from specific demographic subgroups.

The most notable findings are outlined below. More information about the experiments and detailed findings can be found in chapters 1 through 4 of this report. Chapter 5 also provides a summary of the report's main findings and notes which aspects of the experimental conditions will be incorporated into the next NHES administration.

## Leveraging Response Mode Assignment to Increase Response Rates

As part of an effort to better understand how the offered response modes affect NHES response rates, NHES:2019 included a **choice-plus experiment**, in which sample members were offered web, paper, and inbound telephone response options in all screener-phase mailings. The screener is used in NHES to determine eligibility for second-stage substantive surveys. Sample members also were offered a promised incentive for responding by web or by inbound telephone; most were offered \$10, but some were randomly assigned to be offered \$20.

Among all experimental conditions tested in NHES:2019, the choice-plus conditions were the most successful at maximizing both the screener and overall response rates.<sup>3</sup> While the \$20 condition did not increase the final screener response rate compared to the \$10 condition, it did increase response to the earlier screener mailings and the screener response rate by web. This reduced the need for screener-phase nonresponse follow-up mailings and led to increases in the topical and overall response rates.<sup>4</sup> However, the choice-plus conditions also were among the most expensive conditions included in NHES:2019, suggesting that their use for the full sample in future NHES administrations is likely to be cost prohibitive.

---

<sup>1</sup> “Web-push” is a type of sequential mixed-mode survey design. In this type of design, the initial survey contacts request response by web, and later contacts add or switch to more expensive response modes, such as paper questionnaires.

<sup>2</sup> The NHES is a two-phase survey. The first phase is the screener phase, in which sampled households are asked to provide basic information about all household members age 20 or younger. In all households with an eligible child, a child is sampled for the second, topical phase of the survey, during which the household is asked to provide more detailed information about a sampled child in one of two longer topical surveys.

<sup>3</sup> The overall response rate is the product of the screener-phase response rate and the topical-phase response rate.

<sup>4</sup> Increased screener response by web tends to contribute to higher topical response rates because the topical response rate is consistently higher among web screener respondents, who can proceed directly from the screener to the topical survey, than among paper screener respondents, for whom a separate paper topical survey request must be sent in a later mailing package.

## Refining Web-Push Contact Mailing Procedures

For several of the experiments, the primary goal was to refine the screener-phase web-push mailing procedures.

- **Number of advance mailings:** Sample members were randomly assigned to receive no advance mailings, an advance letter only, or two postcards prior to the advance letter—which we call the “advance mailing campaign” condition.<sup>5</sup>
- **Reminder mailing type:** Sample members in some experimental conditions received a reminder postcard after the initial screener package, while sample members in other conditions received a pressure-sealed envelope.<sup>6</sup>
- **FedEx mailing timing:** Sample members were randomly assigned to be sent the second screener package via FedEx, the third screener package via FedEx, or the fourth screener package via FedEx—or to be part of a “modeled FedEx” condition, in which the timing of each address’s FedEx mailing was determined based on its response propensity and its predicted FedEx shipping cost.<sup>7</sup>

Generally, these variations in the web-push mailing procedures did not have a statistically significant effect on the screener response rate. However, some of the treatments—sending an advance letter, sending a pressure-sealed envelope reminder mailing, and sending the second screener package via FedEx—increased response to the earlier screener mailings and the screener response rate by web. This also often led to higher topical and overall response rates for these treatments.

## Attracting Response From Specific Demographic Subgroups

Finally, two experiments included in NHES:2019 attempted to increase response from specific demographic subgroups.

---

<sup>5</sup> The purpose of these postcards was to build familiarity and engagement with NHES. They displayed NHES:2016 findings but did not mention that the address was sampled for NHES:2019.

<sup>6</sup> A pressure-sealed envelope is a letter that has been enclosed and sealed on itself to allow for transmission of sensitive information without an envelope. The pressure sealing includes perforation for easy opening by the recipient. This is a common mailing format for official documentation, such as pay stubs, tax forms, and bills. The NHES:2019 pressure-sealed envelope included the web survey URL and the household’s unique web login credentials, but the reminder postcard did not (because the postcard format did not allow for sufficient protection of this information). Because the mailings sent after these reminders were not equivalent in the two conditions, the analysis for this comparison is limited to the response rate gain after this reminder mailing only.

<sup>7</sup> In the modeled FedEx condition, based on a cost-weighted response propensity score, some sample members were sent the second screener package via FedEx (those with lower FedEx costs and/or lower predicted response propensity) and others were sent the fourth package via FedEx (those with higher FedEx costs and/or higher predicted response propensity).

- **Opt-out screener materials:** All sample members in this condition were sent screener materials that attempted to increase response among households without NHES-eligible children by conveying that the NHES survey request is low burden for households without children.<sup>8</sup>
- **Targeted Spanish-speaking-household materials:** In this condition, likely-Spanish-speaking households were sent tailored materials that were expected to improve response among Spanish-speaking households.<sup>9</sup>

However, neither of these treatments increased the screener or overall response rates among the targeted subgroups. These results speak to the potential challenges of using targeted approaches in address-based samples—both in developing materials that effectively speak to the target population and in accurately reaching the target population when the information available about sampled households prior to data collection is limited or inaccurate.

---

<sup>8</sup> Specifically, the screener letters included additional language to emphasize that the survey is very short for households without children, and the paper screener questionnaire was modified so that the first item asking whether the household had any children was on the front cover of the paper screener questionnaire booklet, allowing households without children to immediately see that they only needed to respond to that item.

<sup>9</sup> For example, the tailored materials included Spanish-first presentation of the bilingual materials; icons on the envelopes emphasizing that the survey is voluntary, about the respondent's community, and trusted; and images of Hispanic families on the cover of the paper screener questionnaire. These materials were developed based on a series of focus groups with Spanish speakers. Likely-Spanish-speaking households were identified in the NHES:2019 sample based on auxiliary data that was available on the sampling frame or appended from publicly available sources. All other households in this condition were sent the standard NHES materials.

# Contents

Chapter 1. Introduction.....	1
1.1    Overview of NHES:2019 Methodological Experiments .....	1
1.2    NHES:2019 Methodology and Key Outcomes .....	5
Chapter 2. Leveraging Response Mode Assignment to Increase Response Rates .....	11
2.1    Choice-Plus Experiment.....	12
2.2    Modeled-Mode Experiment .....	21
Chapter 3. Refining Web-Push Contact Procedures.....	30
3.1    Advance Mailings Experiment .....	30
3.2    Pressure-Sealed Envelope Versus Reminder Postcard .....	36
3.3    FedEx Timing Experiment .....	39
Chapter 4. Attracting Response From Specific Demographic Subgroups.....	51
4.1    Opt-Out Screener Materials Experiment .....	51
4.2    Targeted Spanish-Speaking-Household Materials Experiment .....	58
Chapter 5. Summary and Conclusions .....	70
5.1    Leveraging Response Mode Assignment to Increase Response Rates.....	71
5.2    Refining Web-Push Contact Procedures.....	72
5.3    Attracting Response From Specific Demographic Subgroups.....	75
References .....	78

# List of Exhibits

Exhibit 1.1.	Overview of NHES:2019 methodological goals and experimental conditions .....	3
Exhibit 1.2.	Allocation of NHES:2019 sample to experimental conditions .....	4
Exhibit 2.1.	Choice-plus experiment key facts.....	13
Exhibit 2.2.	Modeled-mode experiment key facts .....	22
Exhibit 2.3.	Auxiliary variables included in model used to identify paper-sensitive households in NHES:2019 modeled-mode experiment.....	24
Exhibit 2.4.	Assigned screener mailing protocol, by experimental condition and paper- sensitivity classification.....	25
Exhibit 3.1.	Advance mailings experiment key facts .....	31
Exhibit 3.2.	Pressure-sealed envelope versus reminder postcard analysis key facts.....	37
Exhibit 3.3.	FedEx timing experiment key facts .....	40
Exhibit 3.4.	Auxiliary variables included in the response propensity and predicted FedEx cost models .....	42
Exhibit 3.5.	Timing of FedEx mailing, by experimental condition and FedEx-priority assignment.....	43
Exhibit 3.6.	Relative cost per response, by data collection phase and FedEx timing condition: 2019 .....	49
Exhibit 4.1.	Opt-out screener materials experiment key facts .....	52
Exhibit 4.2.	Targeted Spanish-speaking-household materials experiment key facts .....	58
Exhibit 4.3.	Icons used on materials sent to likely-Spanish-speaking households in targeted materials condition .....	61
Exhibit 4.4.	Assigned mailing materials, by experimental condition and likely-Spanish- speaking classification.....	62
Exhibit 5.1.	Screener, topical, and overall response rate, by experimental condition.....	70



## List of Figures

Figure 2.1.1.	Final screener response rate and percentage point gain in screener response rate, by experimental condition and response timing: 2019.....	14
Figure 2.1.2.	Screener response rate, by experimental condition and response mode: 2019 .....	16
Figure 2.1.3.	Topical response rate, by questionnaire and experimental condition: 2019 .....	19
Figure 2.1.4.	Overall response rate, by questionnaire and experimental condition: 2019.....	20
Figure 2.2.1.	Final screener response rate and percentage point gain in screener response rate, by experimental condition and response timing: 2019.....	27
Figure 2.2.2.	Overall response rate, by questionnaire and experimental condition: 2019.....	29
Figure 3.1.1.	Final screener response rate and percentage point gain in screener response rate after each mailing, by advance mailing condition and mailing: 2019.....	32
Figure 3.1.2.	Screener response rate, by advance mailing condition and response mode: 2019 .....	34
Figure 3.1.3.	Overall response rate, by questionnaire and advance mailing condition: 2019 .....	36
Figure 3.2.1.	Screener response rate after reminder mailing and percentage point gain in screener response rate after each mailing, by reminder condition: 2019 .....	38
Figure 3.3.1.	Final screener response rate and percentage point gain in screener response rate after each mailing, by FedEx condition and mailing: 2019 .....	45
Figure 3.3.2.	Screener response rate, by FedEx condition and response mode: 2019.....	46
Figure 3.3.3.	Overall response rate, by questionnaire and FedEx condition: 2019.....	48
Figure 4.1.1.	Final screener response rate and percentage point gain in screener response rate after each mailing, by whether household was flagged as having children, opt-out condition, and mailing: 2019.....	54
Figure 4.1.2.	Screener response rate by whether household was flagged as having children, opt-out condition, and response mode: 2019 .....	55
Figure 4.1.3.	Percentage of screener respondents who reported at least one child eligible for the topical surveys, by whether household was flagged as having children, questionnaire, and opt-out condition: 2019 .....	57

Figure 4.2.1.	Final screener response rate and percentage point gain in screener response rate after each mailing, by likely-Spanish-speaking classification, targeted materials condition, and mailing: 2019 .....	63
Figure 4.2.2.	Screener response rate, by likely-Spanish-speaking classification, targeted materials condition, and response mode: 2019 .....	65
Figure 4.2.3.	Observed and self-reported Spanish-speaking status, by likely-Spanish-speaking classification and experimental condition: 2019.....	69

# Chapter 1. Introduction

The 2019 National Household Education Surveys (NHES:2019) represented the first full-scale NHES administration that primarily used a web-push data collection protocol, following a successful baseline, experimental administration of this approach in the 2016 administration of the NHES (NHES:2016). “Web-push” is a type of sequential mixed-mode survey design. In this type of design, the initial survey contacts request response by web, and later contacts add or switch to more expensive response modes, such as paper questionnaires. As part of the NHES:2019 web-push design, sample members were asked to respond by web in initial mailings and were sent a paper questionnaire in later mailings.

## 1.1 Overview of NHES:2019 Methodological Experiments

The NHES has a rich history of conducting methodological experiments to determine the ideal methods to use to maximize response rates, increase representativeness (ideally minimizing nonresponse bias), and maximize efficiency. This report presents the results of several methodological experiments that were embedded in NHES:2019. These experiments were organized around three key themes, with a chapter of this report dedicated to each one.

- **Chapter 2: Understanding how response mode assignment affects response rates.** The results of the NHES:2016 web-push mixed-mode experiment suggest that the web-push approach yields a lower screener-phase response rate compared to a paper-only data collection but much higher topical-phase response rates. How do we leverage response mode assignment to encourage the highest response rates at both the screener and topical phases? Is it effective to offer both web and paper response options concurrently but offer an incentive for response by web? Can we successfully use information available on the sampling frame to identify households that prefer paper and assign only these households to a paper-only protocol?
- **Chapter 3: Refining and updating NHES screener-phase web-push contact procedures based on findings from NHES:2016.** Building on the success of the NHES:2016 web-push mixed-mode experiment, can we continue to refine and improve upon the screener-phase web-push contact procedures used in the NHES? For example, how many and what types of advance mailings should be sent? Is it more effective to send a reminder postcard or a pressure-sealed envelope? What is the ideal timing for FedEx mailings?
- **Chapter 4: Attracting response from specific demographic subgroups.** Can we improve response among specific demographic subgroups? Although some households without children might assume that an education-related survey is not relevant to them, it is still very important for these households to respond to the screener questionnaire. Can we increase response from such households by giving them an opportunity to “opt out” on the first page of the paper screener questionnaire? In addition, Spanish-speaking households tend to have lower NHES response rates than non-Spanish-speaking households. Can we increase response among Spanish-speaking households by providing materials tailored specifically to them?

The overarching goal of this report is to determine which of these experiments were effective at achieving these objectives, which of the tested procedures should be used in future NHES administrations, and which procedures may require further refinement. To this end, the final chapter of the report (chapter 5)

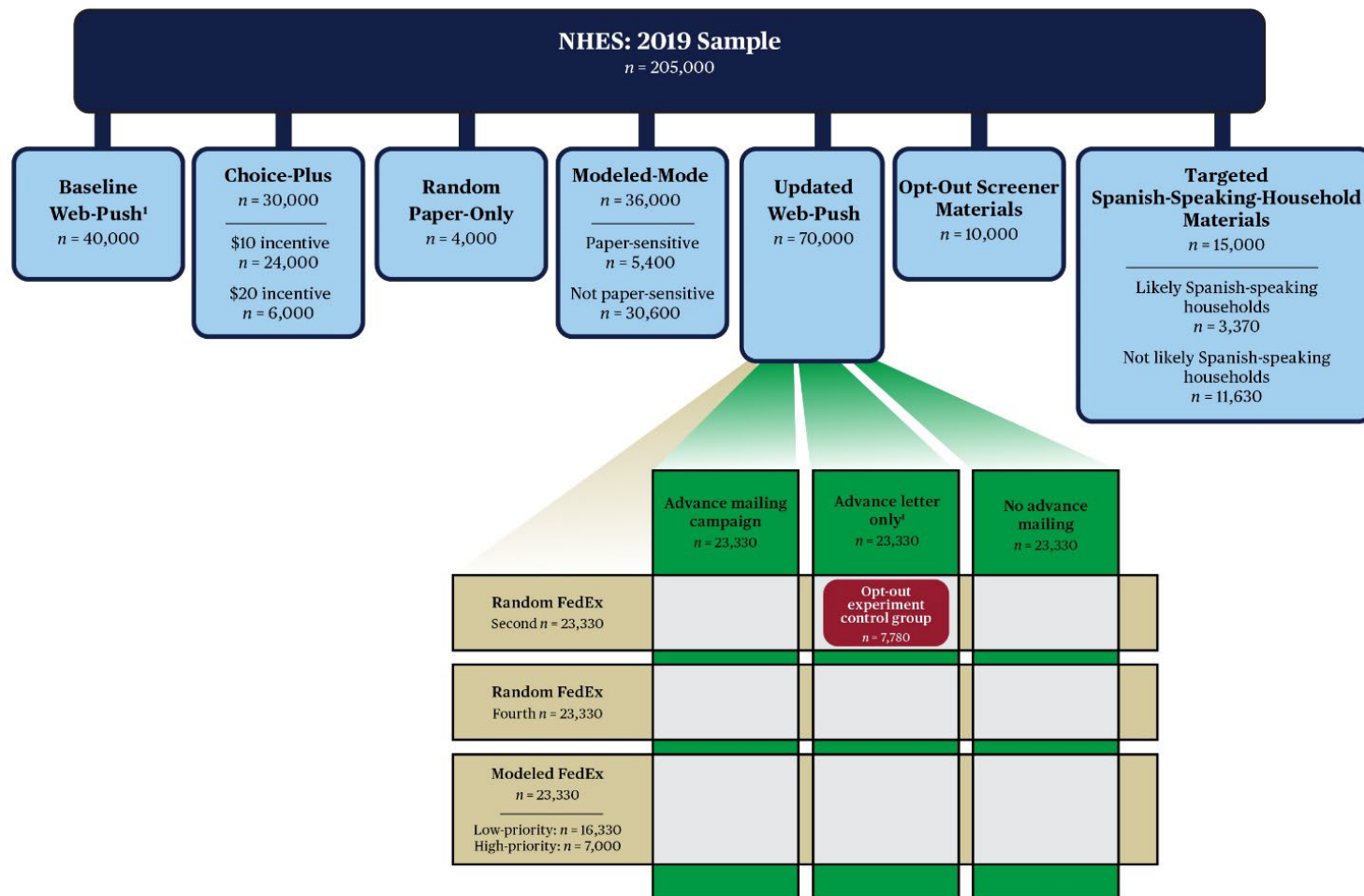
provides overarching conclusions and notes which procedures will be incorporated into the next NHES administration.

Exhibit 1.1 provides an overview of all NHES:2019 methodological goals and experimental conditions, and exhibit 1.2 displays the allocation of the NHES:2019 sample to the experimental conditions. More information about each experiment is provided in later chapters. For most of these experiments, the primary control group was the “baseline web-push condition.” The 40,000 addresses that were randomly assigned to this condition received the same screener mailing protocol used for the NHES:2016 web-push experiment (though with updated contact materials and instruments): a sequential mixed-mode protocol in which the first two screener packages asked for web response and the last two screener packages provided a paper questionnaire. This condition was included to serve as a control group for the experiments that tested enhancements or alternatives to the web-push methodology, as well as to facilitate comparisons of response rates between NHES:2016 and NHES:2019. Some of the NHES:2019 experiments used other conditions as controls in addition to or in lieu of the baseline web-push condition; the control condition(s) for each experiment are identified in the relevant chapter.

**Exhibit 1.1. Overview of NHES:2019 methodological goals and experimental conditions**

Goal	Experimental Condition	Description	Control Condition
<b>Chapter 2:</b> Understand how response mode assignment affects response rates	Choice-plus	In this condition, all mailings offered paper, web, and inbound telephone response options. Sample members were offered a promised incentive if they responded by web or by inbound telephone. Most cases were offered \$10, but some were randomly assigned to be offered \$20.	Baseline web-push Random paper-only
	Modeled-mode	In this condition, “paper-sensitive” cases, which were predicted to prefer paper (based on response patterns to NHES:2016), were only given the option to respond by paper. The remainder of the cases, which were predicted to have lower paper sensitivity, received the baseline web-push protocol.	Baseline web-push Random paper-only
	Random paper-only	All cases assigned to this condition were only given the option to respond by paper.	<i>Not applicable</i>
<b>Chapter 3:</b> Refine screener-phase web-push contact procedures	Updated web-push mailing protocol	Cases assigned to this umbrella condition (“updated web-push condition”) were allocated across two fully crossed experiments that tested refinements to the web-push protocol used in NHES:2016: (1) varying the number advance mailings and (2) varying the timing of FedEx mailings.	
	<i>Advance mailings</i>	Cases were randomly assigned to one of three advance mailing conditions: (1) advance mailing campaign, (2) advance letter only, or (3) no advance mailings.	Advance letter only
	<i>FedEx timing</i>	Cases also were randomly assigned to one of three FedEx conditions: (1) FedEx for the second screener package, (2) FedEx for the fourth screener package, or (3) modeled FedEx (FedEx for either the second or fourth screener package, based on the case’s cost-weighted response propensity score).	Baseline web-push
	<i>Pressure-sealed envelope</i>	All cases in the updated web-push condition received a pressure-sealed envelope after the initial screener package. Those that received an advance letter only were compared to cases in the baseline condition, which received a reminder postcard after the initial screener package.	Baseline web-push
<b>Chapter 4:</b> Attract response from specific demographic subgroups	Opt-out screener materials	All cases in this condition received an “opt-out” paper screener, which had the first item on the questionnaire cover rather than on the inside. The screener letters emphasized that the survey was very short for households without children.	Subset of updated web-push
	Targeted Spanish-speaking materials	In this condition, different materials were used depending on whether a case was identified as being likely-Spanish-speaking. Likely-Spanish-speaking cases within this condition (identified using information on the sampling frame and appended from publicly available sources) received tailored materials that were developed based on the results of focus groups with Spanish-dominant speakers. Not-likely-Spanish-speaking cases within this condition received the baseline web-push materials.	Baseline web-push
Baseline web-push		This condition replicated the web-push design that was tested experimentally in NHES:2016. It served as a control for many of the NHES:2019 experimental conditions.	<i>Not applicable</i>

NOTE: NHES:2019 is National Household Education Surveys Program of 2019. NHES:2016 is National Household Education Surveys Program of 2016.

**Exhibit 1.2.** Allocation of NHES:2019 sample to experimental conditions

<sup>1</sup> The baseline web-push condition also served as the postcard condition. The cases within the updated web-push condition that received an advance letter only also served as the pressure-sealed envelope condition.

NOTE: NHES:2019 is National Household Education Surveys Program of 2019. The baseline web-push condition served as the control condition, except for the following analyses: (1) for the choice-plus and modeled-mode experiments, the random paper-only condition was also a control; (2) for the advance mailings experiment, the advance letter-only condition (part of the updated web-push condition) was the control, (3) for the opt-out screener materials experiment, the control condition was the subset of updated web-push mailing protocol condition that received an advance letter only and FedEx at the second screener package.

## 1.2 NHES:2019 Methodology and Key Outcomes

This section provides an overview of the NHES:2019 methodology and the key data collection outcomes presented in this report. For more information, see the *National Household Education Surveys Program of 2019 Data File User's Manual* (Jackson et al. 2021).<sup>1</sup>

NHES:2019 used an address-based sample covering the 50 states and the District of Columbia. The target population for the survey was all residential addresses in the United States. A sample of 205,000 addresses was drawn from a file of residential addresses that was based on the U.S. Postal Service's Computerized Delivery Sequence File. In addition to address variables, the sampling frame included several auxiliary variables (e.g., gender and age of head of household, presence of children, number of adults in the household) that had been appended to the file by the vendor. These variables were the sole source of information available to the National Center for Education Statistics (NCES) about the household prior to the survey administration; however, it is important to note that they can be subject to missingness and error (Medway et al. 2022).

The survey was administered by the U.S. Census Bureau on behalf of NCES. Data were collected in two phases: the screener phase and the topical phase. At both phases, mail was the primary mode by which sampled addresses were contacted and asked to complete the survey by web, paper, or telephone. Appendix A provides an overview of the screener contact protocol that was used for each experimental condition.

### 1.2.1 Screener-Phase Outcomes

In the first phase of data collection, sampled addresses were asked to complete a screener questionnaire. This questionnaire asked whether there was anyone age 20 or younger living in the household. If so, respondents were asked to provide basic information about each of these household members, such as their name or initials, birth month and year, and school enrollment status. The screener contact materials are provided in appendix B. The paper versions of the screener questionnaire are available in appendix C.

Thus, the first set of key outcomes were related to screener response—whether, when, and in what response mode sampled addresses responded to this screener request, and whether they reported any children who were eligible to be sampled for the topical phase of the survey.

#### 1.2.1.1 Screener Response Rate

The first key screener-phase outcome was the screener response rate. Throughout this report, the screener response rate is defined as the percentage of eligible households that returned a completed screener questionnaire. All screener response rates presented in this report were calculated using

---

<sup>1</sup> The *National Household Education Surveys Program of 2019: Data File User's Manual* is available at <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2021030>.

American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1).<sup>2</sup> Across the entire NHES:2019 administration, the screener response rate was 60 percent.<sup>3</sup>

### **1.2.1.2 Speed of Screener Response**

The second key screener-phase outcome was how early in the screener phase sampled addresses responded to the screener request. Earlier response is desirable due to its ability to increase the efficiency of the data collection. It allows fewer nonresponse follow-up mailings to be sent. For addresses in web-push conditions, it also makes it more likely that the household responded to the screener by web (since the earlier mailings do not include a paper questionnaire).

In all experimental conditions, the screener phase of data collection included several contact attempts. Sample members received either English-only mailings or bilingual contacts.<sup>4</sup> Most mailings were sent via U.S. Postal Service (USPS) First-Class Mail, but one of the nonresponse follow-up packages was sent via FedEx (either the second, third, or fourth screener package, depending on the address's FedEx timing experimental condition). The exception was P.O. box addresses, which cannot receive FedEx mailings and thus were sent the package via U.S. Priority mail instead of FedEx. As a result, 670 P.O. box addresses were excluded from analyses that report on FedEx priority classification and/or timing.

- Most sampled addresses were sent an advance letter (as part of the advance mailings experiment, a small sample of addresses either received additional advance mailings prior to this letter or did not receive an advance letter).
- All sampled addresses were then sent an initial screener package. The package included a cover letter, a \$5 cash incentive, and a Commonly Asked Questions (CAQ) enclosure. Depending on their experimental condition, some addresses also received a paper screener questionnaire and a pre-addressed, postage-paid return envelope.

---

<sup>2</sup> Response Rate 1 (RR1) is calculated as the number of complete responses divided by the sum of the number of complete or partial responses, the number of noncompletes (e.g., refusals), and the number of cases of unknown eligibility (e.g., unknown if housing unit) (AAPOR 2015).

<sup>3</sup> By contrast, the screener response rates reported in the *National Household Education Surveys Program of 2019 Data File User's Manual* (Jackson et al. 2021) are calculated using AAPOR Response Rate 3 (RR3), which incorporates an estimate of the percentage of addresses of unknown eligibility that are eligible into the denominator. Across the NHES:2019 administration, the screener response rate using AAPOR RR3 was 63 percent. However, due to the difficulty of making statistical comparisons between response rates calculated using RR3 because of assumptions needed for allocations of unknown respondent eligibility, all unit response rates presented in this report are calculated using AAPOR RR1, which assumes that all addresses of unknown eligibility status are, in fact, eligible. Therefore, they represent the estimated response rate under the most conservative eligibility assumption and can be interpreted as the proportion of sampled cases (excluding cases known to be ineligible) that returned a completed screener questionnaire.

<sup>4</sup> The bilingual mailings had English on one side of the cover letter and Spanish on the other side, and, when they included a paper questionnaire, included both English and Spanish copies of the questionnaire. The materials were generally displayed in an English-first presentation, with the English materials on top of the Spanish ones. All NHES:2019 screener-phase mailings included a bilingual Commonly Asked Questions enclosure (with English on one side and Spanish on the other), regardless of whether the rest of the mailing was English-only or bilingual. At the topical phase, households were sent either English-only or Spanish-only materials based on the language in which they responded to the screener.



- One week after the initial screener package, either a reminder postcard or a pressure-sealed envelope was sent to each address (depending on its experimental condition). A pressure-sealed envelope is a letter that has been enclosed and sealed on itself to allow for transmission of sensitive information without an envelope. The pressure sealing includes perforation for easy opening by the recipient. This is a common mailing format for official documentation, such as pay stubs, tax forms, and bills. The NHES:2019 pressure-sealed envelope included the web survey URL and the household's unique web login credentials, but the reminder postcard did not (because the postcard format did not allow for sufficient protection of this information).
- Nonresponding addresses were then sent up to three additional nonresponse follow-up mailings (the second, third, and fourth screener packages). The response modes offered in these packages varied based on the address's experimental condition. In all conditions, addresses that had a telephone number available on the sampling frame also received an automated reminder telephone call. This call was made on the same date that the third screener package was mailed; therefore, it was not possible to assess whether sample members responded as a result of receiving this call (as opposed to responding due to receiving the third screener package).

For most experiments, we present the percentage point gain in the screener response rate after each mailing. Response was attributed to a mailing if the response was received 3 or more days after that mailing was sent and less than 3 days after the next mailing was sent. For a smaller number of experiments, we present a more streamlined version of this analysis that focuses on response to the “early mailings” (initial screener package, reminder postcard/pressure-sealed envelope, and second screener package) and the “late mailings” (third screener package and fourth screener package).

### ***1.2.1.3 Mode of Screener Response***

The third key screener-phase outcome was the mode in which the household responded to the screener. The mode of screener response is important because it affects the efficiency of the data collection, as well as having downstream effects on the topical-phase response rates.

Web is the preferred mode of screener response for the NHES. Web responses are received more quickly than paper responses and tend to require less data processing; collecting web responses is also less costly than collecting paper or telephone responses. In addition, when the screener data are submitted via the web instrument, within-household sampling occurs immediately. This allows web respondents to complete both the screener and a topical survey in one sitting, which has a positive effect on the topical-phase response rate. In contrast, respondents to the paper questionnaire must mail the completed screener back to the Census Bureau. If sampled for a topical survey, the household is then sent separate topical-phase mailings that include a prepaid incentive, making paper topical response more costly than web topical response.

As noted earlier, most of the sampled addresses were part of a web-push protocol in which sample members were asked to respond by web in initial mailings and were sent a paper questionnaire in later mailings. As part of the response mode experiments discussed in chapter 2, a smaller number of addresses received web and paper response options for all mailings or only a paper screener questionnaire for all mailings. In all conditions, the contact materials included the Census Bureau's Telephone Questionnaire Assistance (TQA) toll-free number. The choice-plus condition contact materials also mentioned the option to complete the survey over the telephone by using the TQA number. When sample members called the TQA number, they were offered the option to complete the survey over the

telephone with a Census Bureau telephone interviewer; such responses are described as “inbound telephone” responses elsewhere in the report.

Therefore, we present three mode-specific screener response rates in this report.

- **Screener response rate by web:** The percentage of eligible households that submitted a completed screener questionnaire via the web instrument.
- **Screener response rate by paper:** The percentage of eligible households that returned a completed paper screener questionnaire.
- **Screener response rate by inbound telephone:** The percentage of eligible households that completed the screener questionnaire over the telephone.

#### ***1.2.1.4 Subgroup Screener Response Rates***

The fourth key screener-phase outcome was subgroup screener response rates. These response rates were calculated using more than 20 household- and area-level variables that were available on the NHES sampling frame or were appended from publicly available data sources (e.g., the American Community Survey). The effect of the experimental conditions on subgroup response rates is of interest because differential effects across subgroups affect the composition of screener respondents; for example, if a particular experimental condition caused a higher-than-average increase in the response rate for a particular subgroup, it would increase the representation of that subgroup among respondents. The composition of screener respondents is of interest because it can affect the extent of nonresponse bias in the NHES estimates or the variability of nonresponse-adjusted weights and, therefore, the precision of estimates.

#### ***1.2.1.5 Topical Eligibility Rate***

The final key screener-phase outcome was the topical eligibility rate. The topical eligibility rate is of interest because, all else equal, a higher topical eligibility rate will yield a larger topical sample from a given number of screener respondents.

Once completed screener questionnaires were received, the screener responses were used to determine if there were any eligible children living in the household. The topical eligibility rate represents the percentage of screener respondent households that reported a child who was eligible to be sampled for one of the topical-phase surveys. We report three topical eligibility rates throughout this report.

- **Early Childhood Program Participation (ECP) topical survey eligibility rate:** The percentage of screener respondent households that reported at least one child age 6 or younger and not yet enrolled in kindergarten.
- **Parent and Family Involvement in Education (PFI) topical survey eligibility rate:** The percentage of screener respondent households that reported at least one child ages 3-20 and enrolled in public or private school in grades kindergarten through 12 (or homeschooled for the equivalent).
- **Overall topical eligibility rate:** The percentage of screener respondent households that reported at least one child eligible for either the ECP or the PFI.

No more than one child per household was sampled for a topical survey. Because a household could report both ECPP- and PFI-eligible children, the overall topical eligibility rate is less than the sum of the ECPP- and PFI-specific topical eligibility rates.

### **1.2.2 Topical-Phase Outcomes**

As noted above, in the second, topical phase of data collection, screener respondent households that reported at least one child eligible for a topical survey were asked to complete a longer topical questionnaire with questions about the sampled child's care arrangements or education. Paper screener respondents that were sampled for a topical survey were sent an initial topical package requesting that the household complete the included paper topical questionnaire; this package also included a prepaid cash incentive of either \$5 or \$15. Topical cash incentive amounts were determined based on the timeliness of a household's screener response. Households that responded to the screener by paper after the third screener package were considered "late responders" that received the larger \$15 topical-phase incentive. Households that responded to the screener by paper earlier received the \$5 incentive. Nonrespondents received a reminder postcard, up to three additional topical packages, and—when a telephone number was available on the sampling frame—an automated reminder call. In general, web and inbound telephone screener respondents that were sampled for a topical survey did not receive topical-phase mailings because they were automatically routed to the topical survey immediately after completing the screener, and most of them completed the topical survey at that time.<sup>5</sup>

The key topical-phase outcome was the topical response rate, with separate response rates being calculated for the ECPP and PFI topical surveys. These rates were calculated using AAPOR RR1—that is, the percentage of households sampled for the topical survey that ended up completing it. Across the entire NHES:2019 collection, the topical response rates for the ECPP and PFI surveys were 86 percent and 83 percent, respectively.<sup>6</sup>

We also present mode-specific topical response rates in appendix E. These represent the topical response rate among the topical-eligible households that completed the screener in that mode (e.g., the topical response rate by web is the percentage of households that completed the topical survey among those who were sampled for it and completed the screener on the web).

### **1.2.3 Overall Outcomes**

The final key outcome was the overall response rate, with separate response rates again being calculated for the ECPP and PFI topical surveys. The overall response rates were calculated as the product of the screener response rate and the topical response rate for the relevant topical survey. Across the entire NHES:2019 collection, the overall response rates for the ECPP and PFI surveys were 51 percent and

---

<sup>5</sup> The only exception was the small number of web screener respondents that were sampled for a topical survey but chose not to complete any topical survey items immediately after completing the screener. They received a letter encouraging web topical completion along with a cash incentive, a reminder postcard, up to three additional topical packages (the last two of which included a paper questionnaire), and—when a telephone number was available on the frame—an automated reminder call.

<sup>6</sup> These are the same as the topical response rates reported in the *National Household Education Surveys Program of 2019 Data File User's Manual* (Jackson et al. 2021). The topical response rates use AAPOR RR1 because there are no cases of unknown eligibility at the topical phase.

50 percent, respectively.<sup>7</sup> Mode-specific and subgroup-specific overall response rates are presented in appendix E and were calculated using the same methods described earlier in this chapter.

The remainder of this report presents the results of the methodological experiments embedded into NHES:2019. All analyses were conducted using base weights. Statistically significant differences between experimental conditions were identified using *t* tests unless otherwise noted. All discussed effects of experimental conditions can be assumed to be statistically significant unless otherwise noted. We did not adjust for multiple comparisons (e.g., by using a Bonferroni correction). Finally, all in-text references to percentage point differences are based on unrounded estimates. Therefore, readers may find that a difference cited in the text may not be identical to the difference obtained by using the rounded estimates shown in the accompanying figures.

---

<sup>7</sup> By contrast, the overall response rates reported in the *National Household Education Surveys Program of 2019: Data File User's Manual* (Jackson et al. 2021) are 54 percent for the ECPP and 53 percent for the PFI. These differ from the rates reported here because they are based on a screener response rate calculated using AAPOR RR3 rather than AAPOR RR1.

## Chapter 2. Leveraging Response Mode Assignment to Increase Response Rates

The NHES:2016 administration included an experiment to test a web-push design in which sampled households received mailed invitations to complete the survey by web before being sent nonresponse follow-up mailings with paper questionnaires. The results of this experiment suggested that using a web-push design is preferable to paper-only for most NHES sample members—although the screener response rate was slightly lower for the web-push condition than for the paper-only condition, the topical response rates were much higher.<sup>1</sup> The NHES:2016 overall response rates also were higher for the experimental web-push condition than for the paper-only condition. However, based on the lower screener response rate in the web-push condition, it was hypothesized that there may be a cohort of “paper-sensitive” cases that will not respond to the screener unless offered a paper questionnaire from the beginning of the data collection and who may have already decided not to participate by the time they receive a paper questionnaire in later contact attempts.

Thus, despite the conclusion that web-push was generally preferable to paper-only for most NHES sample members, several outstanding questions remained after the NHES:2016 web-push experiment.<sup>2</sup> In general, how do we leverage response mode assignment to encourage the highest response rates at both the screener and topical phases? In particular, would it be more effective to offer both web and paper response options concurrently but offer an incentive for response by web? Alternatively, can we successfully use information available on the sampling frame to identify households that prefer paper and assign only these households to a paper-only protocol?

To address these questions, NHES:2019 tested two potential approaches to reducing the impact of paper sensitivity on response rates while still obtaining web responses from as many cases as possible.

- A **choice-plus** experiment, which used a concurrent mixed-mode design (with web, paper, and inbound telephone response options in all screener packages) and offered a promised incentive for web or inbound telephone response to encourage response in those modes.
- A **modeled-mode** experiment, in which cases that were predicted to be paper-sensitive were assigned to a paper-only protocol while all other cases in that condition received a web-push protocol.

Each section of this chapter begins with a description of the experiment and then discusses the effect of the experimental treatment on key outcomes.

---

<sup>1</sup> Detailed response rates for the NHES:2016 experimental treatment groups can be found in chapter 5 of the *NHES:2016 Data File User's Manual* (McPhee et al. 2018).

<sup>2</sup> Since the NHES is a multi-phase survey, the ideal design is likely to be different for researchers who are conducting single-phase surveys. For single-phase surveys, the NHES:2016 experiment suggests that paper-only may be preferable in terms of maximizing response.

## 2.1 Choice-Plus Experiment

The NHES:2019 choice-plus experiment used a design adapted from an experiment that was included in the 2015 Residential Energy Consumption Survey (RECS) National Pilot Study (Biemer et al. 2018), sponsored by the U.S. Energy Information Administration. In this type of “choice-plus” design, sample members are given a choice of multiple response modes from the start of the data collection, with one or more of the offered modes incentivized over the others. This incentive structure is designed to make the incentivized mode(s)—which are generally the mode(s) preferred by the data collector to maximize cost savings or data quality—more attractive and avoid the phenomenon seen in prior research where offering a concurrent mode choice can result in lower response rates (Dillman, Smyth, and Christian 2014).

The NHES choice-plus experiment differed from the RECS experiment in the following ways: (1) while the RECS choice-plus condition incentivized web response only, the NHES choice-plus condition incentivized response by both web and inbound telephone; (2) while all of the RECS experimental conditions also offered promised incentives for response irrespective of response mode, the NHES did not offer such incentives; and (3) while the RECS promised incentive for web response was \$10 (in addition to the promised incentive that was provided irrespective of response mode), the NHES tested both \$10 and \$20 promised incentives for web or inbound telephone response.

Exhibit 2.1 summarizes the design and results of the NHES:2019 choice-plus experiment. Within the \$10 and \$20 choice-plus conditions, the incentive was mentioned in all screener mailings. To receive the promised cash incentive, sample members had to meet several criteria, including (1) completing the full survey, including the topical survey if they were sampled for one; and (2) completing the survey by the deadline stated in the fourth screener package. The U.S. Census Bureau sent the promised incentives by mail. In the rest of this section, in addition to comparing the \$10 and \$20 choice-plus conditions to each other, we also compare them to the baseline web-push condition to better understand the effect of the experiment relative to the mailing protocols used in the NHES:2016 web-push experiment. The random paper-only experimental condition is also included as a reference point for how choice-plus differs from a paper-only design.

**Exhibit 2.1. Choice-plus experiment key facts**

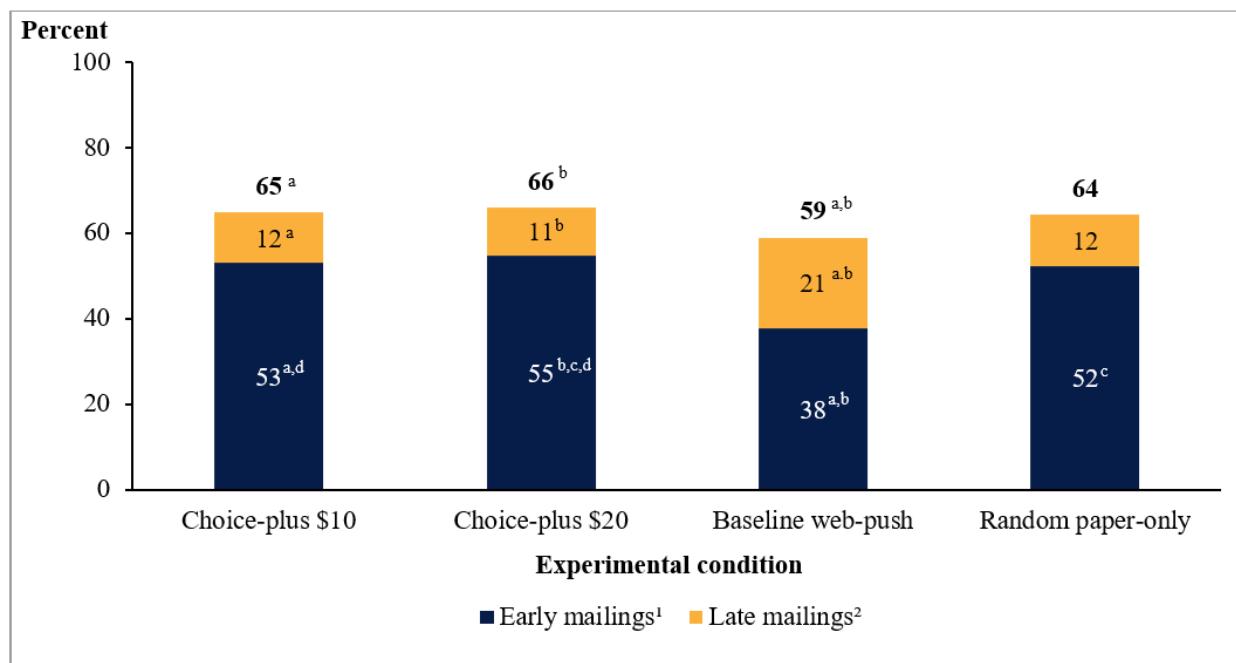
<b>Description</b>	To better understand how the offered response modes affect response, this experiment tested the effectiveness of using a concurrent mixed-mode protocol that uses a promised cash incentive to encourage web and inbound telephone response. This protocol encouraged response in the preferred modes (web and inbound telephone), while still offering a paper questionnaire for those who prefer it.
<b>Treatment conditions</b>	<p><b>Choice-plus, \$10 promised incentive (n=24,000).</b> Cases assigned to this condition received a concurrent mixed-mode protocol in which all screener packages offered web, paper, and inbound telephone response options. They were offered a \$10 promised incentive for responding by web or inbound telephone.</p> <p><b>Choice-plus, \$20 promised incentive (n=6,000).</b> Cases assigned to this condition received the same mailing protocol as the \$10 choice-plus condition, except that they were offered a \$20 promised incentive.</p>
<b>Control conditions</b>	<p><b>Baseline web-push (n=40,000).</b> Cases assigned to this condition received a sequential mixed-mode protocol in which the first two screener packages requested web response and the last two screener packages included a paper questionnaire. They were not offered a promised incentive.</p> <p><b>Random paper-only (n=4,000).</b> Cases assigned to this condition received a paper-only protocol in which all four screener packages only included a paper screener questionnaire. They were not offered a promised incentive.</p>
<b>Case selection</b>	Random assignment to conditions.
<b>Results</b>	<ul style="list-style-type: none"> <li>The ECPP and PFI overall response rates were higher for both choice-plus conditions than for the baseline web-push condition. The overall response rates for the ECPP and PFI also were higher for the \$20 choice-plus condition than for the \$10 choice-plus condition.</li> <li>Both the \$10 and \$20 choice-plus conditions resulted in higher screener response rates than the baseline web-push condition, but there was not a significant difference between the screener response rate for the two incentive levels.</li> <li>The choice-plus conditions yielded higher screener response by paper than the baseline web-push condition, but there was not a significant difference between the two choice-plus conditions and the baseline web-push conditions in screener response by web or inbound telephone. However, the screener response rate by web was significantly higher in the \$20 choice-plus condition than in the \$10 choice-plus condition. Due to the previously established positive association between screener response by web and topical response rates, this also led to higher topical response rates in the \$20 condition than in the \$10 condition.</li> <li>The choice-plus conditions increased screener response among subgroups that are underrepresented in NHES data relative to the baseline web-push condition, suggesting the potential to reduce nonresponse bias.</li> </ul>

NOTE: ECPP is Early Childhood Program Participation topical survey. PFI is Parent and Family Involvement topical survey. NHES is National Household Education Surveys Program.

**2.1.1 Screener Response Rate**

It was hypothesized that the choice-plus condition would yield a higher screener response rate due to the promised incentive, the inclusion of a paper questionnaire in the initial screener package, and the promising results from the RECS experiment. Additionally, the \$20 choice-plus condition was expected to have a higher screener response rate than the \$10 condition.

As shown in figure 2.1.1, **both the \$10 and \$20 choice-plus conditions resulted in higher screener response rates than the baseline web-push condition** (6 and 7 percentage points higher than web-push, respectively; see also table 2.1.1 in appendix E). However, there was not a statistically significant difference between the response rates for the choice-plus conditions and the random paper-only condition at the screener phase. Additionally, contrary to the expected results, **there was not a statistically significant difference between the response rates for the two choice-plus conditions at the screener phase.**

**Figure 2.1.1. Final screener response rate and percentage point gain in screener response rate, by experimental condition and response timing: 2019**

<sup>a</sup> Significant difference ( $p < .05$ ) between choice-plus \$10 condition and baseline web-push condition.

<sup>b</sup> Significant difference ( $p < .05$ ) between choice-plus \$20 condition and baseline web-push condition.

<sup>c</sup> Significant difference ( $p < .05$ ) between choice-plus \$20 condition and random paper-only condition.

<sup>d</sup> Significant difference ( $p < .05$ ) between \$10 and \$20 choice-plus conditions.

<sup>1</sup> The early mailings are the initial screener package, the pressure-sealed envelope (in the choice-plus conditions)/reminder postcard (in the baseline web-push or random paper-only conditions), and the second screener package.

<sup>2</sup> The late mailings are the third screener package and fourth screener package.

NOTE: Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that completed the screener after the specified mailings. Sample members were considered to have responded to the early mailings if their screener response was received no more than 2 days after the third screener package was mailed; sample members were considered to have responded to the late mailings if their screener response was received 3 or more days after the third screener package was sent. Unweighted eligible sample size was 22,040 for the \$10 choice-plus condition, 5,530 for the \$20 choice-plus condition, 36,700 for the baseline web-push condition, and 3,680 for the random paper-only condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

Because choice-plus respondents were offered a promised incentive and a paper questionnaire from the beginning of data collection, it was hypothesized that the effect of this condition would be seen after the early screener mailings.<sup>3</sup> As expected, **the response rates after the early screener mailings for the \$10 and \$20 choice-plus conditions were higher than for the baseline web-push condition** (by 15 percentage points and 17 percentage points, respectively) (see also table 2.1.2 in appendix E).

<sup>3</sup> The “early mailings” are the initial screener package, the pressure-sealed envelope (in the choice-plus conditions)/reminder postcard (in the baseline web-push or random paper-only conditions), and the second screener package. Sample members were considered to have responded to the early mailings if their screener response was received no more than 2 days after the third screener package was mailed.



The response rate after the early mailings also was higher in the \$20 condition than in the random paper-only condition, though the magnitude of the difference was quite small (2 percentage points). One limitation of this analysis is that the choice-plus condition used a pressure-sealed envelope for the reminder mailing, while the baseline web-push and random paper-only conditions used a postcard. Therefore, to better understand how useful the pressure-sealed envelope was in the choice-plus condition, we compared the effect of the pressure-sealed envelope in the updated web-push mailing protocol condition to the effect of the postcard in the baseline web-push condition (see section 3.2 for more details). Overall, the pressure-sealed envelope yielded a response rate increase 4 percentage points higher than that of the postcard. Therefore, some of the difference between the choice-plus, baseline web, and random paper-only conditions after the second reminder mailing can be attributed to this different reminder type.

The response rate after the early mailings also was higher in the \$20 choice-plus condition than in the \$10 choice-plus condition; however, the magnitude of the difference was again relatively small (2 percentage points). This suggests that offering the larger promised incentive may slightly increase response to earlier mailings, even though there was not a statistically significant difference in the final screener response rate between the \$10 and \$20 conditions. For the NHES design, an increase in earlier screener response, regardless of final screener response rates, provides the benefit of reducing postage and printing costs associated with nonresponse follow-up mailings.

For the late mailings,<sup>4</sup> the pattern in response gains in the choice-plus conditions relative to the baseline web-push condition was reversed; the gain in response was larger for the baseline web-push condition than for both the \$10 and \$20 choice-plus conditions (by 10 percentage points). These results show that the switch to offering paper in the third and fourth screener packages for the baseline web-push condition helps to make up for some, but not all, of the difference in the early response rates between the baseline web-push and choice-plus conditions. There was not a statistically or practically significant difference in the screener response rate gain after the late mailings between the \$10 choice-plus, \$20 choice-plus, and random paper-only conditions.

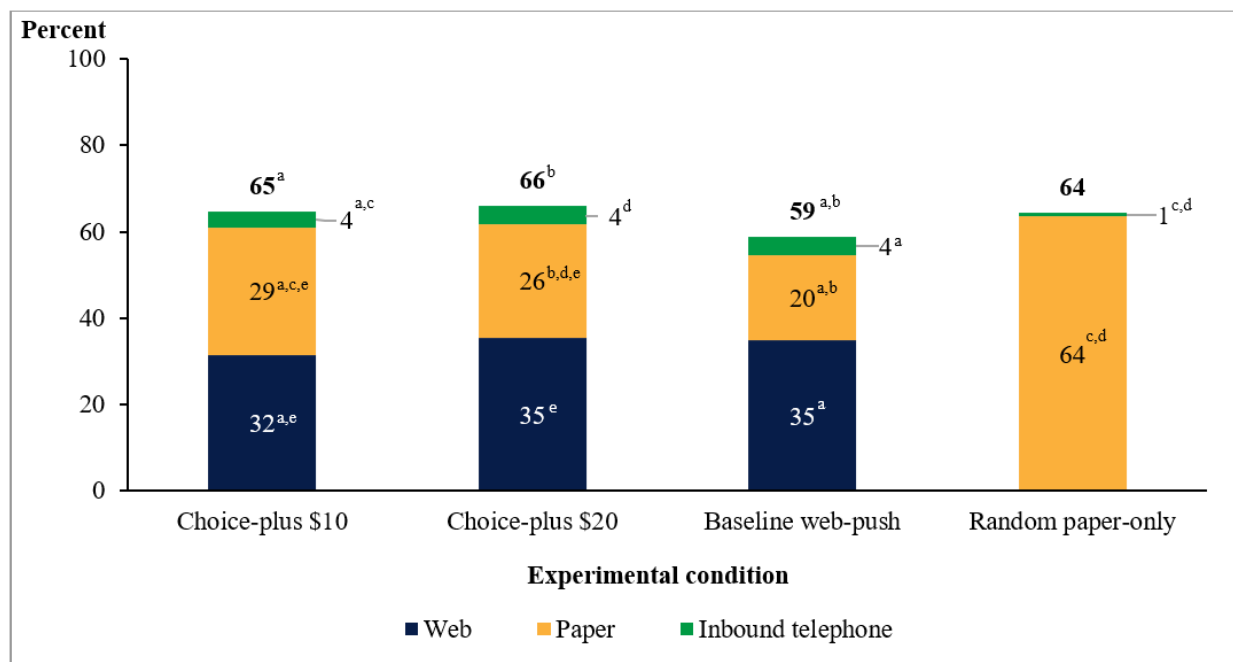
### 2.1.2 Screener Mode of Response

Sample members in the baseline web-push condition did not have an opportunity to respond to the early screener mailings by paper unless they contacted the Census Bureau to ask for a paper questionnaire. So, more paper screener response after the early mailings in the choice-plus conditions than in the baseline web-push condition was a given. However, providing a promised incentive for response by web and inbound telephone might lessen the total proportion of paper responses to *all* screener mailings in choice-plus relative to the baseline web-push condition.

As shown in figure 2.1.2, **the screener response rate by paper was higher in the \$10 and \$20 choice-plus conditions than in the baseline web-push condition, and the size of this difference was smaller when a \$20 incentive was offered (7 percentage points) than when a \$10 incentive was offered (10 percentage points)** (see also table 2.1.1 in appendix E).

---

<sup>4</sup> The late mailings are the third screener package and fourth screener package. Sample members were considered to have responded to the late mailings if their screener response was received 3 or more days after the third screener package was sent.

**Figure 2.1.2. Screener response rate, by experimental condition and response mode: 2019**

<sup>a</sup> Significant difference ( $p < .05$ ) between choice-plus \$10 condition and baseline web-push condition.

<sup>b</sup> Significant difference ( $p < .05$ ) between choice-plus \$20 condition and baseline web-push condition.

<sup>c</sup> Significant difference ( $p < .05$ ) between choice-plus \$10 condition and random paper-only condition.

<sup>d</sup> Significant difference ( $p < .05$ ) between choice-plus \$20 condition and random paper-only condition.

<sup>e</sup> Significant difference ( $p < .05$ ) between \$10 and \$20 choice-plus conditions.

NOTE: The random paper-only condition did not offer a web response option. Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that were respondents to the screener questionnaire. For response rates by inbound telephone, the unrounded estimates are 3.84 percent for the choice-plus \$10 condition, 4.18 percent for the choice-plus \$20 condition, 4.32 percent for the baseline web-push condition, and 0.71 percent for the random paper-only condition. With large sample sizes and small underlying percentages, very small percentage point differences may be statistically significant. Unweighted eligible sample size was 22,040 for the \$10 choice-plus condition, 5,530 for the \$20 choice-plus condition, 36,700 for the baseline web-push condition, and 3,680 for the random paper-only condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

Nevertheless, paper responses still accounted for less than half of the screener response rate in the choice-plus conditions, and the screener response rate by paper remained substantially lower than the random paper-only condition rate (by over 30 percentage points).

However, incentivizing web and inbound telephone responses did not increase the screener response rate by either of these modes in the choice-plus conditions compared to the baseline web-push condition. In fact, the **screener response rate by web was about 3 percentage points lower in the \$10 choice-plus condition than it was in the baseline web-push condition. The screener response rate by web was approximately the same in the \$20 choice-plus condition as in the baseline web-push condition.**

The screener response rate by inbound telephone did not differ meaningfully in any of these three conditions, though the random paper-only condition yielded a statistically significantly lower screener

response rate by inbound telephone than the choice-plus or baseline web-push conditions (by 3 to 4 percentage points).

### 2.1.3 Subgroup Screener Response Rates

The choice-plus conditions led to statistically significant increases in the screener response rate for most subgroups relative to the baseline web-push condition (see table 2.1.3 in appendix E). **There were some subgroups for which the choice-plus conditions led to a larger-than-average increase in the screener response rate as compared to the baseline web-push condition.** These were generally subgroups that had lower-than-average response rates under the baseline web-push condition, implying that the choice-plus conditions improved representativeness. Examples of this pattern can be seen for the variables shown below. For ease of discussion, the percentage point differences listed below are for the \$20 choice-plus condition only. Similar significant differences were observed for the \$10 condition, although the magnitude of the differences tended to be slightly smaller in the \$10 condition than in the \$20 condition.

- **Age of head of household:** For households whose head of household was 18-34 or 35-44 years old, the \$20 choice-plus condition increased the screener response rate by approximately 9 to 10 percentage points, compared to 6 percentage points or less for other age subgroups.
- **Race/ethnicity of head of household:** For households whose head of household was Black, the \$20 choice-plus condition increased the screener response rate by approximately 11 percentage points, compared to 7 percentage points or less for other race/ethnicity subgroups.
- **Household income:** For households whose household income was less than \$25,000, the \$20 choice-plus condition increased the screener response rate by 11 percentage points, compared to 7 percentage points or less for other household income subgroups.
- **Home tenure:** For households that were renter occupied, the \$20 choice-plus condition increased the screener response rate by 11 percentage points, compared to 5 percentage points for households that were owner inhabited.
- **Census Low Response Score:**<sup>5</sup> For households in block groups with the highest Low Response Scores (fourth quartile; that is, those with the lowest mail return rates to the 2010 Decennial Census), the \$20 choice-plus condition increased the screener response rate by 11 percentage points, compared to 7 percentage points for households in block groups with the lowest Low Response Scores (first quartile; that is, those with the *highest* mail return rates to the 2010 Decennial Census).

Overall, **there was not much variation in subgroup response rates between the \$20 incentive and the \$10 incentive choice-plus conditions.** Two exceptions, where the \$20 incentive had a statistically significant increase in response compared to \$10, were for households where the head of household had

---

<sup>5</sup> The Census Low Response Score is a derived variable that identifies block groups with characteristics associated with low mail return rates to the 2010 Decennial Census. The variable is available from the Census Planning Database at <https://www.census.gov/topics/research/guidance/planning-databases.html>. A higher Low Response Score corresponds to a lower expected mail return rate.

a bachelor's degree as the highest level of education and households whose household income was \$150,000 or more (about 5 percentage points for both). However, these subgroups already had higher-than-average response rates, even with a \$10 incentive. For most subgroups, the screener response rates for the choice-plus conditions were not significantly different from the random paper-only condition.

#### 2.1.4 Topical Eligibility Rate

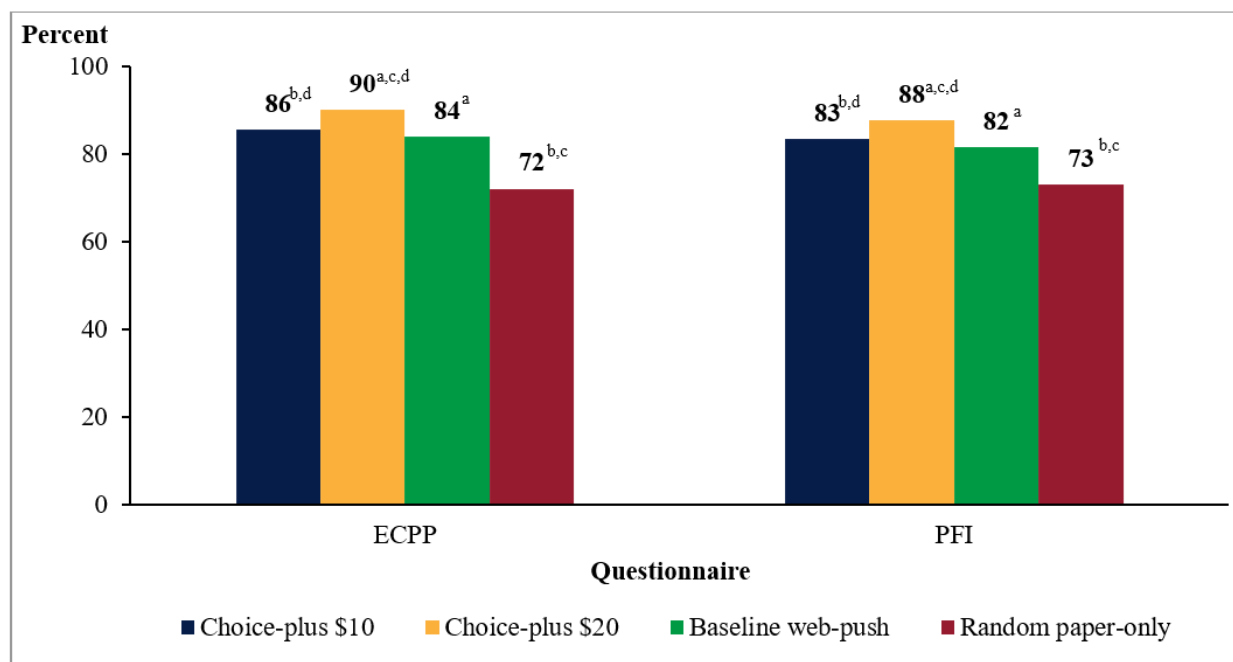
**There was not a significant difference between the topical eligibility rate for the choice-plus conditions (regardless of incentive level) and the baseline web-push condition.** The incentive value in the choice-plus conditions also did not have a significant effect on the topical eligibility rate (see table 2.1.4 in appendix E). However, the topical eligibility rate was higher among choice-plus respondents (regardless of incentive level) than among random paper-only respondents (by about 2 percentage points).

#### 2.1.5 Topical Response Rates and Overall Response Rates

It was expected that the effect of the choice-plus condition on the topical and overall response rates would be dependent on its effect on screener response rates and on the mode of screener response. For example, a higher screener response rate by web would help increase the topical response rates (which could, in turn, increase the overall response rates); however, increased screener response by paper would have the potential to reduce the topical response rates because a separate topical questionnaire would need to be mailed to eligible households, making it necessary to re-engage the household (which could, in turn, reduce the overall response rates).

As shown in figure 2.1.3, **for both the ECPP and the PFI, the \$20 choice-plus condition yielded a higher topical response rate than the \$10 condition** (by 5 and 4 percentage points, respectively) (see also table 2.1.5 in appendix E). This was likely driven by the higher percentage of screener respondents responding by web in the \$20 condition than in the \$10 condition; the topical response rate tends to be much higher for web screener respondents than for paper screener respondents.

Furthermore, the \$20 choice-plus condition yielded statistically significantly higher topical response rates than the baseline web-push condition (by 6 percentage points for both surveys), whereas the \$10 choice-plus condition did not. Finally, as expected, the topical response rates for both surveys were higher for both choice-plus conditions than for the random paper-only condition.

**Figure 2.1.3. Topical response rate, by questionnaire and experimental condition: 2019**

<sup>a</sup> Significant difference ( $p < .05$ ) between choice-plus \$20 and baseline web-push condition.

<sup>b</sup> Significant difference ( $p < .05$ ) between choice-plus \$10 and random paper-only condition.

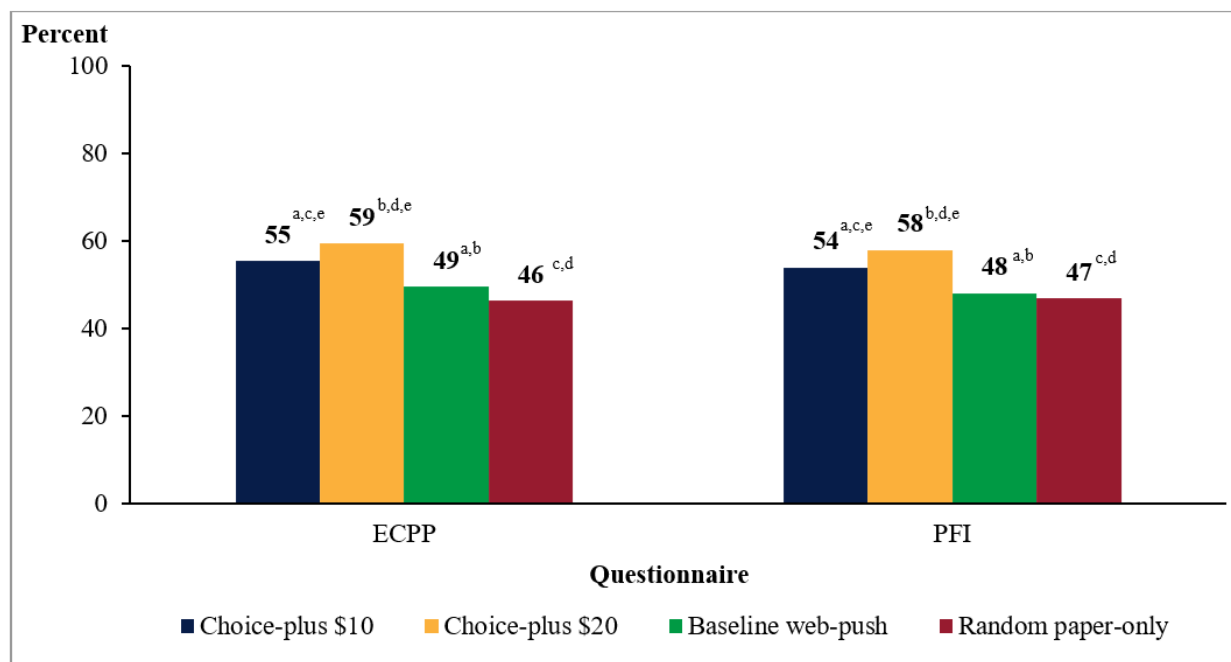
<sup>c</sup> Significant difference ( $p < .05$ ) between choice-plus \$20 and random paper-only condition.

<sup>d</sup> Significant difference ( $p < .05$ ) between \$10 and \$20 choice-plus conditions.

NOTE: ECPP is Early Childhood Program Participation topical survey. PFI is Parent and Family Involvement topical survey. Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Percentages represent the proportion of eligible households that were respondents to the topical questionnaire. Unweighted eligible sample size for the ECPP response rate was 1,140 for the \$10 choice-plus condition, 290 for the \$20 choice-plus condition, 1,620 for the baseline web-push condition, and 150 for the random paper-only condition. Unweighted eligible sample size for the PFI response rate was 2,570 for the \$10 choice-plus condition, 670 for the \$20 choice-plus condition, 3,740 for the baseline web-push condition, and 410 for the random paper-only condition. Sample sizes have been rounded to the nearest 10.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

As shown in figure 2.1.4, for both the ECPP and the PFI, the overall response rate was higher in the \$20 choice-plus condition than in the \$10 choice-plus condition (by about 4 percentage points for each topical survey). This was mostly driven by increased overall response by web in the \$20 condition versus the \$10 condition (see also table 2.1.6 in appendix E). In addition, the overall response rates for both surveys were higher for the \$10 and \$20 choice-plus conditions than for the baseline web-push condition (by 6 and 10 percentage points, respectively). This was mostly driven by increased response by paper in the choice-plus conditions compared to the baseline web-push condition. Finally, the overall response rates for the choice-plus condition also were higher than the overall response rates for the random paper-only condition (by 7 to 9 percentage points for the \$10 choice-plus condition and by 11 to 13 percentage points for the \$20 choice-plus condition).

**Figure 2.1.4. Overall response rate, by questionnaire and experimental condition: 2019**

<sup>a</sup> Significant difference ( $p < .05$ ) between choice-plus \$10 and baseline web-push condition.

<sup>b</sup> Significant difference ( $p < .05$ ) between choice-plus \$20 and baseline web-push condition.

<sup>c</sup> Significant difference ( $p < .05$ ) between choice-plus \$10 and random paper-only condition.

<sup>d</sup> Significant difference ( $p < .05$ ) between choice-plus \$20 and random paper-only condition.

<sup>e</sup> Significant difference ( $p < .05$ ) between \$10 and \$20 choice-plus conditions.

NOTE: ECPP is Early Childhood Program Participation topical survey. PFI is Parent and Family Involvement topical survey. The ECPP and PFI overall response rates are the product of the screener response rate and the respective topical response rate. Screener and topical response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Unweighted eligible sample size for the ECPP response rate was 22,040 for the \$10 choice-plus condition, 5,530 for the \$20 choice-plus condition, 36,700 for the baseline web-push condition, and 3,680 for the random paper-only condition. Unweighted eligible sample size for the PFI response rate was 22,040 for the \$10 choice-plus condition, 5,530 for the \$20 choice-plus condition, 36,700 for the baseline web-push condition, and 3,680 for the random paper-only condition. Sample sizes have been rounded to the nearest 10.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

For reference, tables 2.1.7 and 2.1.8 in appendix E show the ECPP and PFI overall response rates, respectively, broken out by various subgroups of interest. When there were statistically significant differences, they tended to be in the direction of higher response rates in the choice-plus conditions than in the baseline web-push and random paper-only conditions. However, these comparisons should be

interpreted with caution due to small sample sizes, particularly for the \$20 choice-plus and random paper-only conditions.<sup>6</sup>

## 2.2 Modeled-Mode Experiment

Considering the lower screener response rate by web-push than by paper-only in the NHES:2016 web-push experiment and bearing in mind that not all U.S. households have internet connectivity, it was hypothesized that there may be households for whom only sending a paper questionnaire from the start of the data collection would be more effective at gaining response than using a sequential mixed-mode web-push protocol in which a paper questionnaire is not sent until the later part of the data collection.

To identify these sample members, we operationalized measurement of a cohort of “paper-sensitive” sample members—that is, those households that are more likely to respond to the screener when offered a paper questionnaire from the *beginning* of the data collection than they are to respond under a web-push protocol. In theory, if these “paper-sensitive” cases could be identified prior to data collection, they could be offered a paper questionnaire in all screener packages, and this would increase the screener response rate. However, a potential tradeoff is that, by increasing the percentage of screener questionnaires that are completed by paper, this approach could also reduce the topical response rate.

The purpose of the modeled-mode experiment was to (1) determine whether paper-sensitive cases could be accurately identified prior to data collection, and (2) assess whether—if such cases were provided with a paper questionnaire from the beginning of the collection—the resulting screener response rate increase would be sufficient to offset the potential topical response rate decrease (relative to what would have been attained had all cases, regardless of paper sensitivity, received a web-push screener protocol). Exhibit 2.2 summarizes the design and results of the experiment.

---

<sup>6</sup> In the NHES, the theoretical formula for the variance of the overall response rate incorporates the variances of both its individual components—the screener and topical response rates—as well as some complex covariance terms. In practice, replicate weights incorporate all of these individual sources of variation. However, if a cell has a very high topical response rate in the sample, there will be very little variation in that rate across the jackknife replicates. At the extreme, if everyone sampled for the topical survey within a cell responded, then the topical response rate would be 100 percent in all replicates, the estimate of the variance would be 0 for the topical phase, and thus the estimate of the variance of the overall response rate would be equal to the variance of the screener response rate. But, in a small cell, that estimate of a 100 percent topical response rate is likely incorrect, as it is very unlikely to be the case that everyone in that subgroup, if sampled, would respond. The true topical variance is probably non-0, but we cannot estimate it because there is no variation across the replicates. As a result, the standard error of the overall response rate is likely to be underestimated for small subgroups.

**Exhibit 2.2. Modeled-mode experiment key facts**

<b>Description</b>	<p>In an effort to better understand how response mode assignment affects response rates, the purpose of this experiment was to test the effectiveness of using a paper-only protocol for those cases expected to be most likely to prefer responding by paper.</p> <p>As part of this experiment, a predictive model was used to identify <i>paper-sensitive cases</i>—the 15 percent of cases most likely to prefer to respond by paper—based on response patterns to NHES:2016. All other cases were considered <i>non-paper-sensitive cases</i> (those who do not have as much of a preference to respond by paper).</p>
<b>Treatment condition</b>	<b>Modeled-mode (n=36,000).</b> Paper-sensitive cases in this condition (n=5,400) received a paper-only protocol in which all four screener packages contained a paper questionnaire only. Non-paper-sensitive cases in this condition (n=30,600) received the same protocol as the baseline web-push condition.
<b>Control conditions</b>	<p><b>Baseline web-push (n=40,000).</b> Regardless of their paper-sensitivity classification, cases assigned to this condition received a web-push protocol in which the first two screener packages requested web response and the last two screener packages included a paper questionnaire.</p> <p><b>Random paper-only (n=4,000).</b> Regardless of their paper-sensitivity classification, all cases assigned to this condition received a paper-only protocol in which all four screener packages offered a paper questionnaire.</p>
<b>Case selection</b>	Random assignment to conditions. Modeled assignment to paper-only or web-push protocols within the modeled-mode condition.
<b>Results</b>	<ul style="list-style-type: none"> <li>The modeled-mode condition resulted in a slightly higher screener response rate than the baseline web-push condition but a lower screener response rate than the random paper-only condition.</li> <li>The PFI overall response rate also was slightly higher for the modeled-mode condition than for the baseline web-push condition.</li> <li>For both paper-sensitive and non-paper-sensitive cases, the screener response rate was higher under a paper-only protocol than a web-push protocol. Immediately after the second screener package, the response rate difference was larger in the paper-sensitive cohort than in the non-paper-sensitive cohort; this suggests that the predictive model successfully identified cases whose response to early mailings was sensitive to the initially offered mode. However, this was no longer the case by the end of the screener field period; this suggests that the model did not identify cases whose final screener response status was sensitive to the full sequence of offered modes.</li> </ul>

NOTE: NHES:2016 is National Household Education Surveys Program of 2016. PFI is Parent and Family Involvement topical survey.

**2.2.1 Identifying Paper-Sensitive Cases**

Several steps, detailed in the following sections, were taken to identify paper-sensitive cases.

**2.2.1.1 Developing a sensitivity model with NHES:2016 data**

In the modeled-mode experiment, “paper sensitivity” was defined as the change in a case’s response propensity after the first two screener packages that could be attributed to the use of a paper-only protocol instead of a web-push protocol. Using NHES:2016 data for which response outcomes were already known, a logistic regression model was used to estimate this difference (for more information about the performance of this model, please see Jackson et al. 2021).

**Dependent variable.** Screener response status after the first two screener packages (before the switch to paper in the web-push protocol).



- Although both screener and topical response are relevant in assessing the success of a given data collection protocol in the NHES, the sensitivity model used screener response alone as the dependent variable. This is because the detection of interactions between the assigned protocol and the auxiliary variables requires the protocol to have been randomly assigned. In the NHES:2016 web-push experiment, the mode sequence was assigned to a random sample at the screener phase; at the topical phase, the mode depended on the mode of screener response and therefore was no longer randomly assigned.
- The dependent variable was specified as response to either of the first two screener packages because these are the mailings at which web responses were requested within the web-push protocol; it was hypothesized that any variations in preference for paper versus web would be most apparent in response to these screener packages. Furthermore, preliminary analysis found this specification to generate more accurate predictions of paper sensitivity than a specification that used response to any of the four screener packages as the dependent variable.

**Independent variables.** Included (1) auxiliary variables that were available on the sampling frame or appended from publicly available sources (exhibit 2.3),<sup>7</sup> (2) a treatment indicator for whether the case was randomly assigned to the web-push protocol (indicator=1) or the paper-only protocol (indicator=0), and (3) interactions between these auxiliary variables and the treatment indicator. The interactions allowed the estimated impact of the assigned data collection protocol to vary by the available auxiliary variables, allowing the model to be used to predict response propensity under each potential protocol (web-push or paper-only).

---

<sup>7</sup> A conditional random forest model was used to select a subset of the 47 candidate covariates for inclusion in the model.

**Exhibit 2.3. Auxiliary variables included in model used to identify paper-sensitive households in NHES:2019 modeled-mode experiment**

Variable	Source
Address flagged as having Hispanic surname	Address-level characteristic from sampling frame
Person ages 18 to 24 present at address	Address-level characteristic from sampling frame
Person ages 35 to 64 present at address	Address-level characteristic from sampling frame
Person ages 65 or over present at address	Address-level characteristic from sampling frame
Age of head of household	Address-level characteristic from sampling frame
Census region	Address-level characteristic from sampling frame
Dwelling type	Address-level characteristic from sampling frame
Education of head of household	Address-level characteristic from sampling frame
Race/ethnicity of head of household	Address-level characteristic from sampling frame
Gender of head of household	Address-level characteristic from sampling frame
Household income	Address-level characteristic from sampling frame
Marital status of head of household	Address-level characteristic from sampling frame
Number of adults in household	Address-level characteristic from sampling frame
Number of children in household	Address-level characteristic from sampling frame
Home tenure	Address-level characteristic from sampling frame
Availability of telephone number	Address-level characteristic from sampling frame
Residential Internet penetration	Tract-level estimate from FCC
Low Response Score	Block group-level estimate from ACS
Median household income	Block group-level estimate from ACS
Percent Hispanic persons	Block group-level estimate from ACS
Percent married couples	Block group-level estimate from ACS
Percent Black persons	Block group-level estimate from ACS
Percent persons without high school diploma	Block group-level estimate from ACS
Percent persons speaking non-English language	Block group-level estimate from ACS
Percent persons ages 25 to 44	Block group-level estimate from ACS
Percent persons ages 65 and over	Block group-level estimate from ACS
Percent renters	Block group-level estimate from ACS
Percent occupied housing units	Block group-level estimate from ACS
Tract-level poverty rate	Tract-level estimate from ACS
Race/ethnicity stratum	Tract-level estimate from ACS

NOTE: NHES is National Household Education Surveys Program. FCC is Federal Communications Commission. ACS is the U.S. Census Bureau's American Community Survey.

### 2.2.1.2 Applying the model to NHES:2019 cases to create paper-sensitivity scores

This sensitivity model developed with the NHES:2016 data was then applied to the NHES:2019 sample. Two response propensity scores were estimated for each NHES:2019 case:

- (1) the response propensity after the first two screener packages for a web-push protocol ( $\hat{\rho}_w$ ; estimated by setting the treatment indicator in the model to 1); and
- (2) the response propensity after the first two screener packages for a paper-only protocol ( $\hat{\rho}_p$ ; estimated by setting the treatment indicator in the model to 0).

A paper-sensitivity score was estimated for each case by subtracting its response propensity under the web-push protocol ( $\hat{\rho}_w$ ) from its response propensity under the paper-only protocol ( $\hat{\rho}_p$ ).

### 2.2.1.3 Classifying NHES:2019 cases as paper-sensitive or non-paper-sensitive

The paper-sensitivity scores were then used to classify all sample cases as either paper-sensitive (cases with higher paper-sensitivity scores—that is, those with a larger difference in their response propensity between the web-push and paper-only protocols) or non-paper-sensitive (cases with lower paper-sensitivity scores—that is, those with a smaller difference in their response propensity between the web-push and paper-only protocols). More specifically, the 15 percent of cases with the highest paper-sensitivity scores were classified as the “paper-sensitive” group. Because the paper-sensitivity scores were greater than 0 for nearly all cases, the paper-sensitive cohort consisted of cases for which paper-only caused a particularly large increase in predicted response propensity. As shown in appendix D, analysis of the model showed that it was successful in differentiating paper-sensitive and non-paper-sensitive cases.

### 2.2.1.4 Determining cases’ mode protocol based on their paper-sensitivity classification

In the rest of this section, the modeled-mode condition is compared to two conditions: the baseline web-push condition and the random paper-only condition.

- As shown in exhibit 2.4, among the cases assigned to the **modeled-mode condition**, those in the paper-sensitive group received only paper questionnaires for all screener and topical mailings. The rest of the cases assigned to the modeled-mode condition, the “non-paper-sensitive” cases, received the same mailings as the baseline web-push condition.
- In the baseline **web-push condition**, all cases (regardless of paper sensitivity) received a sequential mixed mode web-push protocol.
- In the **random paper-only condition**, all cases (regardless of paper sensitivity) received only a paper questionnaire in all four screener packages.

**Exhibit 2.4. Assigned screener mailing protocol, by experimental condition and paper-sensitivity classification**

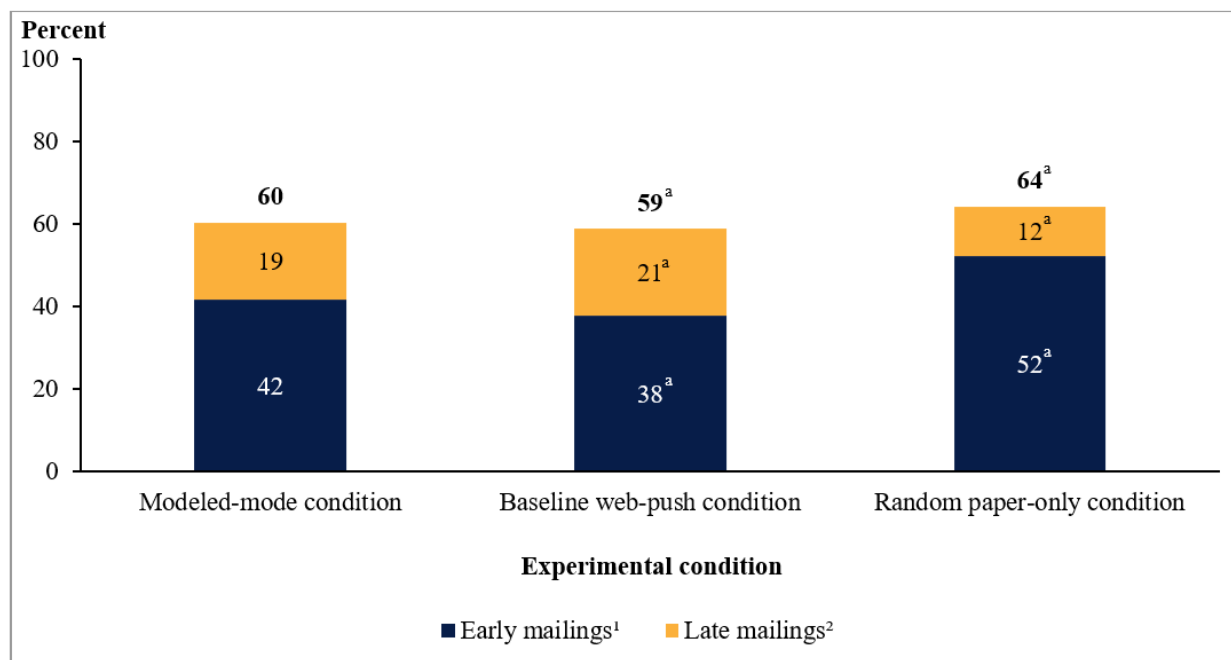
Paper-sensitivity classification	Experimental condition		
	Modeled-mode	Baseline web-push	Random paper-only
Paper-sensitive	<b>Paper-only:</b> All four screener packages include paper questionnaire only	<b>Web-push:</b> <ul style="list-style-type: none"> <li>• Initial and second screener packages request web response</li> <li>• Third and fourth screener packages include paper questionnaire only</li> </ul>	<b>Paper-only:</b> All four screener packages include paper questionnaire only
Non-paper-sensitive	<b>Web-push:</b> <ul style="list-style-type: none"> <li>• Initial and second screener packages request web response</li> <li>• Third and fourth screener packages include paper questionnaire only</li> </ul>		

The analyses in the remainder of this section address this question: if a paper-only protocol is used for the 15 percent of the sample with the highest modeled paper sensitivity and a web-push protocol for the remaining cases (as was done in the modeled-mode condition), how does this affect screener and overall response rates, relative to using a single protocol (web-push or paper-only) for all cases?

### 2.2.2 Screener Response Rate

It was expected that, because the modeled-mode condition provided some cases with a paper questionnaire starting with the initial screener package, it would lead to a higher screener response rate than the baseline web-push condition (which did not provide a paper questionnaire to any cases until the third screener package). Additionally, the random paper-only condition was expected to lead to the highest screener response rate overall.

As shown in figure 2.2.1, consistent with expectations, **the modeled-mode condition resulted in a screener response rate that was slightly higher than in the baseline web-push condition** (by almost 2 percentage points) **but lower than in the random paper-only condition** (by almost 4 percentage points) (see also tables 2.2.1 and 2.2.2 in appendix E). Differences in the early screener response rate were larger in magnitude but similar in direction: the early screener response rate under the modeled-mode condition was about 4 percentage points higher than in the baseline web-push condition and 10 percentage points lower than in the random paper-only condition. Thus, **relative to the use of web-push for all cases, the use of a model to determine the mode of response to the early mailings encouraged somewhat earlier screener response but ultimately led to only a small increase in the final screener response rate.**

**Figure 2.2.1. Final screener response rate and percentage point gain in screener response rate, by experimental condition and response timing: 2019**

<sup>a</sup> Significant difference ( $p < .05$ ) from modeled-mode condition.

<sup>1</sup> The early mailings are the initial screener package, reminder postcard, and second screener package.

<sup>2</sup> The late mailings are the third screener package and fourth screener package.

NOTE: Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RRI). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that completed the screener after the specified mailings. Sample members were considered to have responded to the early mailings if their screener response was received no more than 2 days after the third screener package was mailed; sample members were considered to have responded to the late mailings if their screener response was received 3 or more days after the third screener package was sent. Unweighted eligible sample size was 32,970 for the modeled-mode condition, 36,700 for the baseline web-push condition, and 3,680 for the random paper-only condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

### 2.2.3 Subgroup Screener Response Rates

Table 2.2.3 in appendix E shows the effect of the modeled-mode condition on the screener response rate broken out by various subgroups of interest. There were **no subgroups for which the modeled-mode condition resulted in a higher screener response rate than the random paper-only condition**. There were, however, a few groups for which the modeled-mode condition resulted in a screener response rate relatively close to that achieved for the random paper-only condition (less than 2 percentage points lower). These include households where the head of household was over 65 years old, households with a Black head of household, and households without a telephone number available on the sampling frame.

Some of these patterns likely reflect the ways in which paper-sensitive cases—which received the paper-only protocol under the modeled-mode condition—differ from non-paper-sensitive cases. Because paper-sensitive cases were the only ones that received paper-only mailings under the modeled-mode condition,

the effect of the modeled-mode condition on the screener response rate (relative to the baseline web-push condition) was, by definition, concentrated in this cohort. For example, households with a head of household over 65 years old were more likely than average to be classified as paper-sensitive and therefore more likely than average to receive paper-only mailings under the modeled-mode condition.

To reduce the risk of nonresponse bias, it would be ideal to increase the response rate within a cohort that has lower-than-average response rates under the baseline approach (in this case, a web-push protocol). In practice, however, the opposite was the case; as shown in table 2.2.1 in appendix E, the paper-sensitive cases had a *higher* screener response rate than the non-paper-sensitive cases under the web-push protocol. For example, addresses whose head of household was over 65 years old were more likely to respond in general and more likely to be paper-sensitive; sending them a paper-only protocol in the modeled-mode condition increased their response rate instead of increasing the response rate among a lower-propensity or underrepresented group. This suggests that **the modeled-mode condition did not reduce the risk of nonresponse bias, because the increase in the screener response rate was concentrated among cases that already had higher-than-average response rates.**

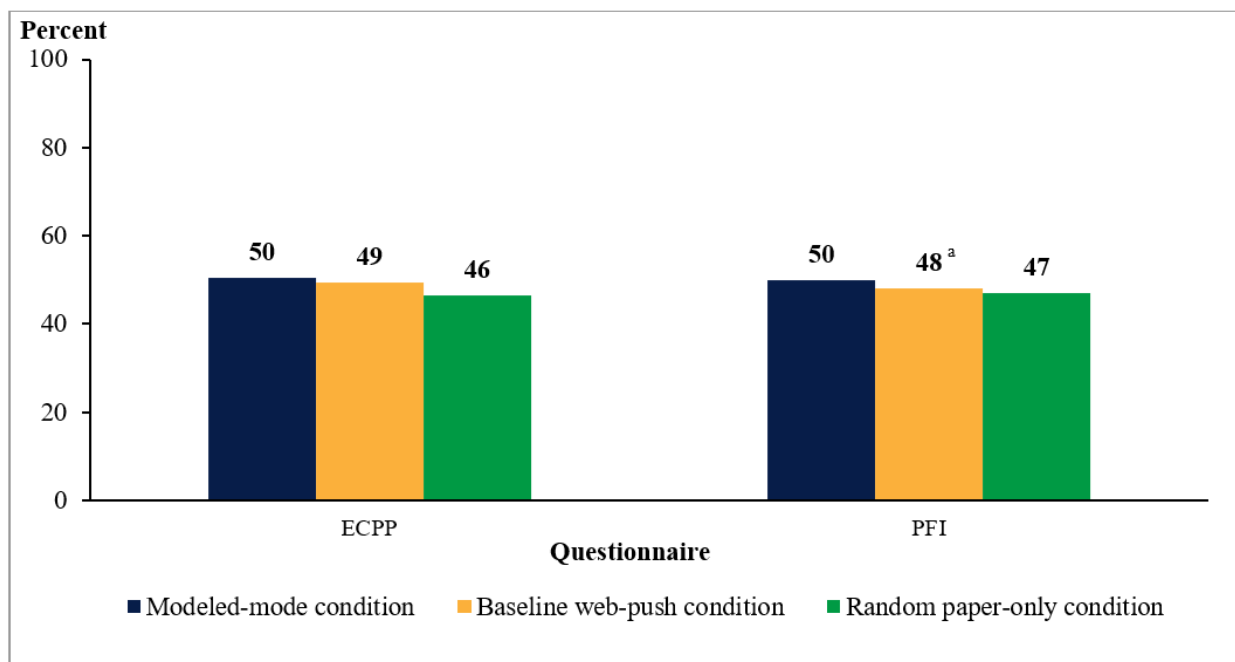
#### 2.2.4 Topical Eligibility Rate

Based on the variables that predicted paper sensitivity in the predictive model, it was anticipated that the random paper-only and modeled-mode conditions would draw more older, childless households to participate and could lead to a lower topical eligibility rate than was achieved in the baseline web-push condition. As shown in table 2.2.4 of appendix E, paper-sensitive cases did, in fact, have much lower topical eligibility rates (by 18-19 percentage points) than non-paper-sensitive cases. Ultimately, however, **the topical eligibility rate was not significantly different between the modeled-mode, baseline web-push, and random-paper only conditions** (within 2 percentage points of each other). Therefore, the response mode sequence had no impact on the percentage of screener respondent households that reported containing NHES-eligible children.

#### 2.2.5 Topical Response Rates and Overall Response Rates

In the NHES:2016 web-push experiment, cases that responded to the screener by paper had lower topical response rates than those that responded to the screener by web. Therefore, it was expected that the modeled-mode condition, by increasing the proportion of screeners completed by paper, would reduce topical response rates relative to the baseline web-push condition, potentially offsetting some of the gain in the screener response rate. This tradeoff between screener and topical response is captured by the overall response rate, the product of the screener and topical response rate.

As shown in figure 2.2.2, **the overall PFI response rate was 2 percentage points higher for the modeled-mode condition than for the baseline web-push condition** (see also table 2.2.6 in appendix E). The overall ECPP response rate did not differ significantly between the modeled-mode and baseline web-push conditions. The overall ECPP and PFI response rates also did not differ significantly between the modeled-mode condition and the random paper-only condition, which may be due to small sample sizes at the topical level.

**Figure 2.2.2. Overall response rate, by questionnaire and experimental condition: 2019**

<sup>a</sup> Significant difference ( $p < .05$ ) from modeled-mode condition.

NOTE: ECPP is Early Childhood Program Participation topical survey. PFI is Parent and Family Involvement topical survey. The ECPP and PFI overall response rates are the product of the screener response rate and the respective topical response rate. Screener and topical response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Unweighted eligible sample size for the ECPP response rate was 32,970 for the modeled-mode condition, 36,700 for the baseline web-push condition, and 3,680 for the random paper-only condition. Unweighted eligible sample size for the PFI response rate was 32,970 for the modeled-mode condition, 36,700 for the baseline web-push condition, and 3,680 for the random paper-only condition. Sample sizes have been rounded to the nearest 10.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

Thus, the impact of the modeled-mode condition on overall response rates was similar to the impact on the screener response rate—namely, a small increase relative to the baseline web-push condition. This is because the ECPP and PFI topical response rates under the modeled-mode condition were approximately the same as under the baseline web-push condition (see table 2.2.5 in appendix E). This may reflect the fact that, as noted above, paper-sensitive cases were much less likely to report an ECPP- or PFI-eligible child on the screener; this implies that a substantial portion of the cases who were only offered paper under the modeled-mode condition were not asked to respond to a topical survey, limiting the impact on the topical response rate.

Again, however, it is noteworthy that paper-sensitive cases already had higher-than-average overall response rates under the web-push protocol (see table 2.2.6 in appendix E). Thus, while the modeled-mode condition slightly increased overall response rates, there is no evidence that it succeeded in bringing in cases that would have been underrepresented under the baseline web-push condition.

For reference, tables 2.2.7 and 2.2.8 in appendix E show the ECPP and PFI overall response rates, respectively, broken out by various subgroups of interest. These comparisons should be interpreted with caution due to small sample sizes, particularly for the random paper-only condition.

## Chapter 3. Refining Web-Push Contact Procedures

The results of the NHES:2016 web-push experiment suggested that a web-push design was preferable to a paper-only design for the NHES moving forward. Therefore, web-push was the design used predominantly in NHES:2019, and a subset of the sample was randomly assigned to receive the identical mailing protocol as that used in the NHES:2016 web-push condition. However, further testing was needed to refine and improve upon the web-push contact procedures used in the NHES. For example, how many and what types of advance mailings should be sent? Is it more effective to send a reminder postcard or a pressure-sealed envelope? What is the ideal timing for FedEx mailings? Which procedures increase the screener response rate by web?

Therefore, to improve upon the success of the NHES:2016 web-push experiment, a subset of the NHES:2019 sample was randomly assigned to an updated web-push mailing protocol whose goal was to further refine the NHES web-push methodology. Specifically, the updated web-push mailing protocol was used in several randomly assigned conditions that, together, implemented the following experiments:<sup>1</sup>

- **The advance mailings experiment**, which aimed to determine the number and types of advance mailings to send.
- **The pressure-sealed envelope experiment**, which aimed to compare the effectiveness of a pressure-sealed envelope versus a reminder postcard.
- **The FedEx timing experiment**, which aimed to determine the ideal timing for sending reminder mailings via FedEx.

The advance mailings and FedEx timing experiments were fully crossed (that is, each condition of the advance mailings experiment occurred in combination with each condition of the FedEx timing experiment); however, this report focuses on the main effect of each experiment because the experiments were not designed to have sufficient sample sizes to examine interactions between the experimental treatments. The effects of the two experiments were not expected to interact, but they were crossed to maximize the efficiency of the design. Readers should be aware that, in each condition of the advance mailing and FedEx timing experiments, one-third of the cases were assigned to each of the conditions in the other experiment.

Each of these experiments is discussed further in the following sections. Each section begins with a description of the experiment and then presents the effect of the experiment on key outcomes.

### 3.1 Advance Mailings Experiment

In NHES:2016, to maintain comparability with the paper-only condition, the web-push experimental condition utilized an advance letter that was sent prior to the initial screener package. The goals of the

---

<sup>1</sup> The analyses in this chapter exclude the updated web-push mailing protocol cases that were assigned to the opt-out screener materials condition, which are discussed in chapter 4.



advance mailings experiment in NHES:2019 were to determine whether an effective web-push design involves sending an advance letter and whether an advance mailing campaign is more effective than an advance letter at increasing response. Sample members receiving the updated web-push mailing protocol were randomly assigned to one of three advance mailing conditions:

- **An advance letter-only condition**, in which households were sent a letter a few days before the initial screener package letting them know that the survey invitation was coming soon.
- **An advance mailing campaign condition**, in which households were sent two oversized, glossy postcards prior to the advance letter (for a total of three advance mailings). These postcards presented interesting statistics from NHES:2016, but they did not mention that the household had been sampled for NHES:2019. It was hypothesized that these mailings would increase familiarity with and build engagement with the NHES.
- **A no-advance mailings condition**, in which households were not sent any mailings prior to the initial screener package. The purpose of this condition was to confirm that there was a benefit to sending an advance letter in the context of an NHES web-push design. Although this had been tested experimentally in a paper-only design during a 2011 field test, it had not been tested in a web-push design.

The advance mailings can be found in appendix B. Exhibit 3.1 summarizes the design and results of the advance mailings experiment.

### Exhibit 3.1. Advance mailings experiment key facts

<b>Description</b>	As part of an effort to refine the web-push methodology that was first tested in NHES:2016, the purpose of this experiment was to determine the ideal number and types of advance mailings to send prior to the survey invitation in a web-push protocol. All conditions in this experiment were part of the updated web-push mailing protocol.
<b>Treatment conditions</b>	<p><b>Advance mailing campaign (n=23,330).</b> Cases assigned to this condition were sent two oversized postcards highlighting interesting statistics that resulted from NHES:2016. The postcards did not indicate that the household had been sampled for NHES:2019. Following the postcards, these cases also were sent an advance letter.</p> <p><b>No advance mailings (n=23,330).</b> Cases assigned to this condition did not receive any advance mailings.</p>
<b>Control condition</b>	<b>Advance letter only (n=23,330).</b> Cases assigned to this condition were sent an advance letter prior to the initial screener package.
<b>Case selection</b>	Random assignment to conditions.
<b>Results</b>	<ul style="list-style-type: none"> <li>• Sending an advance letter increased response to the initial screener package and slightly increased the screener response rate by web. It also led to a small increase in the final screener response rate.</li> <li>• There was no additional gain in screener or overall response from the advance mailing campaign compared to only sending an advance letter.</li> </ul>

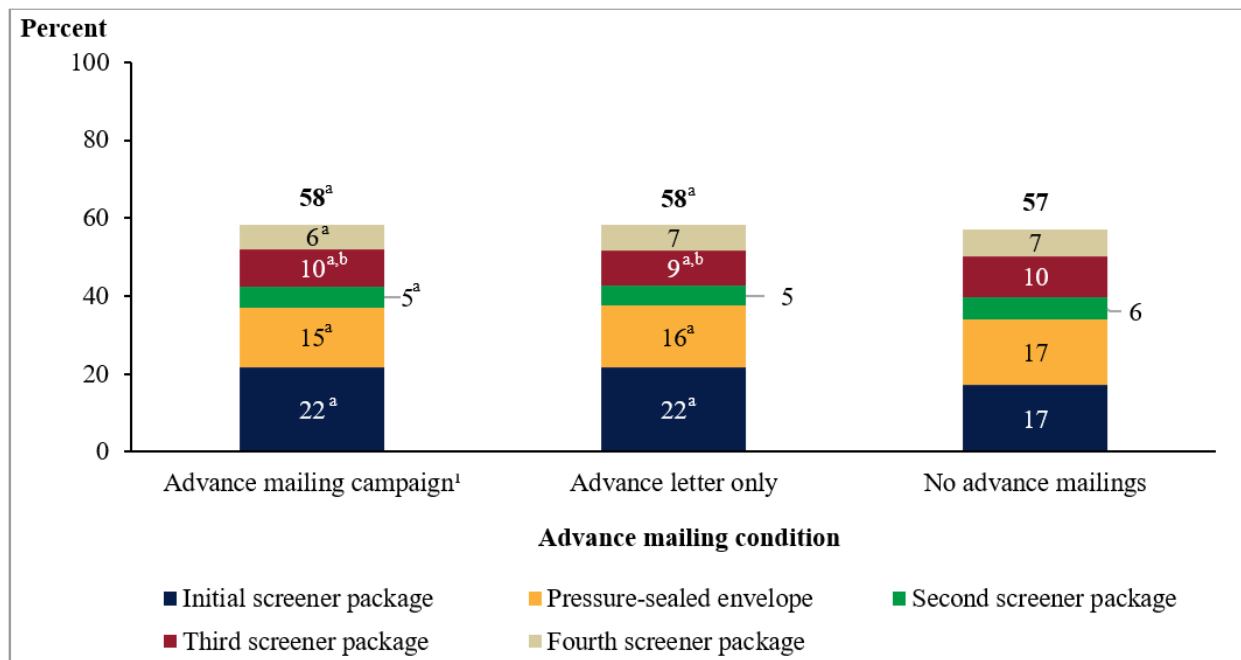
NOTE: NHES:2016 is National Household Education Surveys Program of 2016. NHES:2019 is National Household Education Surveys Program of 2019.

### 3.1.1 Screener Response Rate

It was hypothesized that the advance letter would increase the screener response rate compared to no advance letter—and that the advance mailing campaign would increase the screener response rate compared to the other two conditions. As shown in figure 3.1.1, the **screener response rate was about 1 percentage point higher in the advance letter-only and advance mailing campaign conditions**

than it was in the no-advance mailings condition, a small but statistically significant increase (see also table 3.1.1 in appendix E). However, there was **no difference between the screener response rate in the advance mailing campaign condition and the advance letter-only condition**, suggesting that sending two additional postcards prior to the advance letter offered no additional benefit for the final screener response rate.

**Figure 3.1.1. Final screener response rate and percentage point gain in screener response rate after each mailing, by advance mailing condition and mailing: 2019**



<sup>a</sup> Significant difference ( $p < .05$ ) from no-advance mailing condition.

<sup>b</sup> Significant difference ( $p < .05$ ) between advance mailing campaign and advance letter-only conditions.

<sup>1</sup> Cases in the advance mailing campaign condition received two glossy postcards, followed by an advance letter. These postcards included interesting findings from NHES:2016 but did not mention that the address had been sampled for NHES:2019.

NOTE: NHES:2016 is National Household Education Surveys Program of 2016. NHES:2019 is National Household Education Surveys Program of 2019. Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that completed the screener after the specified mailing. Response is attributed to a mailing if the response was received 3 or more days after that mailing was sent and less than 3 days after the next mailing was sent. All conditions in this experiment were part of the updated web-push condition. Unweighted sample size was 21,230 for the advance mailing campaign condition, 21,260 for the advance letter-only condition, and 21,410 for the no-advance mailings condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

It was also hypothesized that sending advance mailings would increase *early* response to the screener—and that this would be particularly true for the advance mailing campaign. Households that were sent an advance letter or were part of the advance mailing campaign condition responded to the survey earlier than those in the no-advance mailing condition. **The response rate after the initial screener package was 4 percentage points higher in the advance mailing conditions than it was in the no-advance**

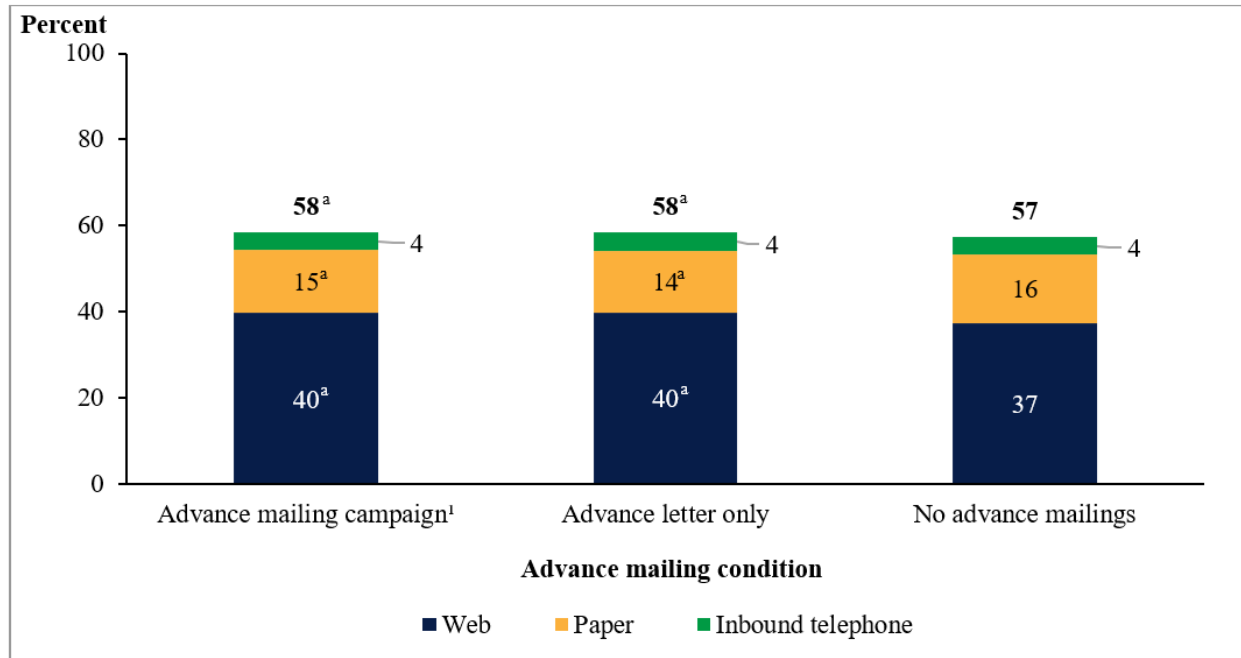
**mailing condition** (see also table 3.1.2 in appendix E).<sup>2</sup> The gain from the advance mailings was much smaller for subsequent mailings, with the no-advance mailing condition doing as well or even slightly better than the two advance mailing conditions for all subsequent mailings; while some of the subsequent differences were statistically significant, they were small in magnitude and were not practically meaningful (most differences were less than 1 percentage point). This result suggests that most of the gain in the screener response rate obtained from sending advance mailings was achieved after the initial screener package. The advance mailing campaign condition did not increase the response to the initial screener package compared to the advance letter-only condition.

### 3.1.2 Screener Mode of Response

It was hypothesized that, because of increased earlier response, the advance mailing conditions would have more response by web than would the no-advance mailings condition—and that this would be particularly true for the advance mailing campaign condition. As shown in figure 3.1.2, **the screener response rate by web was 2 to 3 percentage points higher in the advance mailing conditions than it was in the no-advance mailings condition** (see also table 3.1.1 in appendix E). This was due to the advance letter encouraging response to the earlier mailings; the bulk of the increase in response to the initial screener package was due to increased web response. However, **the screener response rate by paper was about 2 percentage points higher in the no-advance mailings condition than in the advance letter-only condition**. Again, this was driven by households in the no-advance mailing condition responding to later mailings; the bulk of the increase in response to the third screener package was due to increased paper response. The screener response rates in each mode were equivalent in the advance mailing campaign and advance letter-only conditions.

---

<sup>2</sup> Throughout section 3.1, “the advance mailing conditions” refers to both the advance letter-only condition and the advance mailing campaign condition.

**Figure 3.1.2. Screener response rate, by advance mailing condition and response mode: 2019**

<sup>a</sup> Significant difference ( $p < .05$ ) from no-advance mailing condition.

<sup>1</sup> Cases in the advance mailing campaign condition received two glossy postcards, followed by an advance letter. These postcards included interesting findings from NHES:2016 but did not mention that the address had been sampled for NHES:2019.

NOTE: NHES:2016 is National Household Education Surveys Program of 2016. NHES:2019 is National Household Education Surveys Program of 2019. Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that were respondents to the screener questionnaire. All conditions in this experiment were part of the updated web-push condition. Unweighted eligible sample size was 21,230 for the advance mailing campaign condition, 21,260 for the advance letter-only condition, and 21,410 for the no-advance mailings condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

### 3.1.3 Subgroup Screener Response Rates

Table 3.1.3 in appendix E shows the effect of the advance mailing conditions on the screener response rate broken out by various subgroups of interest. Looking at subgroup screener response rates helps understand whether the mailings were particularly effective (or ineffective) for any subgroups.

There was not a statistically significant or practically meaningful difference between the screener response rate for the advance mailing campaign and advance letter-only conditions for any of the subgroups. Moreover, for most of the subgroups, the difference in the screener response rate between each of the advance mailing conditions and the no-advance mailing condition was very small.

The most notable exception was that **the advance mailings appeared to be particularly effective at encouraging response from Black households**. For example, the screener response rate among households that had a Black head of household was 3 to 4 percentage points higher when one or more advance mailings were sent than when no advance mailings were sent; a similar pattern was observed for households in the Black oversample stratum of the NHES sample design (i.e., Census tracts with

25 percent or more Black persons) and households in block groups with the highest percentage of persons who are Black (i.e., in the fourth quartile). Because Black households tend to have a lower screener response rate, this suggests the potential for the advance mailings to improve representativeness.

Finally, an exception to the general trend was that the response rate among households whose head of household was ages 18-34 was about 5 percentage points lower when an advance letter was sent than when no advance mailings were sent; the response rate in the advance mailing campaign condition split the difference between the other two conditions. Because young sample members also tend to have a lower screener response rate, this suggests the potential for the advance letter to reduce representativeness.

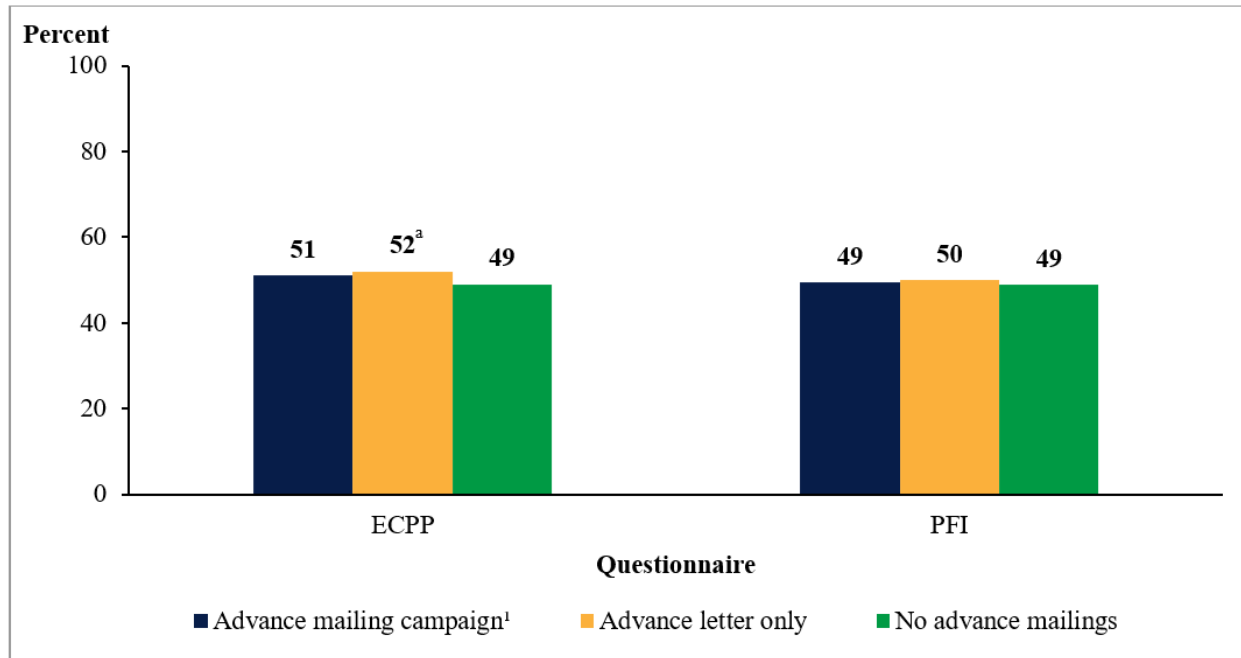
### 3.1.4 Topical Eligibility Rate

The advance mailings were not hypothesized to affect the topical eligibility rate. **The advance mailings did not have an impact on the topical eligibility rate overall or for the PFI in particular** (see table 3.1.4 in appendix E). The topical eligibility rate for the ECPP was statistically significantly higher in the advance mailing campaign condition than in the no-advance mailings condition; however, the magnitude of the difference was less than 1 percentage point.

### 3.1.5 Topical Response Rates and Overall Response Rates

The advance mailings were hypothesized to primarily affect screener-phase response. However, to the extent that the advance mailings encouraged earlier response to the screener, and thus increased web response to the screener, it was hypothesized that they might also have a positive downstream effect on the topical and overall response rates. Although the advance mailings did encourage earlier screener response, they did not have a statistically significant impact on the ECPP and PFI topical response rates (see table 3.1.5 in appendix E).

As shown in figure 3.1.3, for the ECPP, the overall response rate was higher in the advance letter-only condition than in the no-advance mailings condition; however, there was not a statistically significant difference between the advance mailing campaign condition and the no-advance mailings condition (see also table 3.1.6 in appendix E). The advance mailing conditions did not affect the PFI overall response rate. The PFI overall response rate by web was 3 percentage points higher in the advance mailing conditions, but this was offset by a 2-percentage-point decrease in the overall response rate by paper.

**Figure 3.1.3. Overall response rate, by questionnaire and advance mailing condition: 2019**

<sup>a</sup> Significant difference ( $p < .05$ ) from no-advance mailing condition.

<sup>1</sup> Cases in the advance mailing campaign condition received two glossy postcards, followed by an advance letter. These postcards included interesting findings from NHES:2016 but did not mention that the address had been sampled for NHES:2019.

NOTE: NHES:2016 is National Household Education Surveys Program of 2016. NHES:2019 is National Household Education Surveys Program of 2019. ECPP is Early Childhood Program Participation topical survey. PFI is Parent and Family Involvement topical survey. The ECPP and PFI overall response rates are the product of the screener response rates and the respective topical response rate. Screener and topical response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). All conditions in this experiment were part of the updated web-push condition. Unweighted sample size for the ECPP was 21,230 for the advance mailing campaign condition, 21,260 for the advance letter-only condition, and 21,410 for the no-advance mailings condition. Unweighted sample size for the PFI was 21,230 for the advance mailing campaign condition, 21,260 for the advance letter-only condition, and 21,410 for the no-advance mailings condition. Sample sizes have been rounded to the nearest 10.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

Tables 3.1.7 and 3.1.8 in appendix E show the ECPP and PFI overall response rates broken out by various subgroups of interest. As observed for the screener response rate, for nearly all subgroups, there was not a statistically significant or meaningful improvement in the overall response rates in the advance mailing campaign condition compared to the advance letter-only condition. While a handful of subgroups showed statistically significant differences, they tended not to be consistent between the PFI and ECPP. These comparisons should be interpreted with caution, as the sample for this experiment was not designed for subgroup analyses at the topical level.

### 3.2 Pressure-Sealed Envelope Versus Reminder Postcard

In NHES:2016, to maintain comparability with the paper-only condition, the web-push experimental condition sent a reminder postcard after the initial screener package. Moving forward with web-push as the principal NHES design, there was interest in transitioning to a pressure-sealed envelope. The key benefit of the pressure-sealed envelope is that, unlike a postcard, the sealed format is secure enough to

permit including the web survey URL and the household's unique web login credentials. Therefore, while the postcard reminds the sample member to respond without providing an immediate way to do so, the pressure-sealed envelope gives the sample member what they need to complete the web survey. The results of the 2017 NHES web test suggested a pressure-sealed envelope could be a promising option (Medway, Noel, and Guarino 2018); however, this was a web-only administration and thus not completely comparable to the web-push design. In addition, a reminder postcard and pressure-sealed envelope had not been tested simultaneously within the same NHES data collection.

In NHES:2019, cases randomly assigned to the baseline web-push condition were sent a reminder postcard, and cases randomly assigned to the updated web-push mailing protocol were sent a pressure-sealed envelope (see appendix B for the materials). As discussed in section 3.1.1, the updated web-push mailing protocol also included an advance mailings experiment. To maximize comparability with the baseline web-push condition, this analysis is limited to the random subset of cases in the updated web-push mailing protocol that were sent an advance letter (that is, it excludes those cases that were part of the advance mailing campaign condition or the no-advance mailing condition). The goal of analyzing response data related to the reminder postcard and pressure-sealed envelope is to determine which is more effective at encouraging screener response—in particular, screener response by web. Because the mailings sent after these reminders were not equivalent in the two conditions, it is not possible to determine their relative effectiveness on the final screener response rate, topical response rates, or overall response rates. This analysis focuses instead on the amount of screener response received in response to this reminder only.

Exhibit 3.2 summarizes the pressure-sealed envelope versus reminder postcard comparison.

### Exhibit 3.2. Pressure-sealed envelope versus reminder postcard analysis key facts

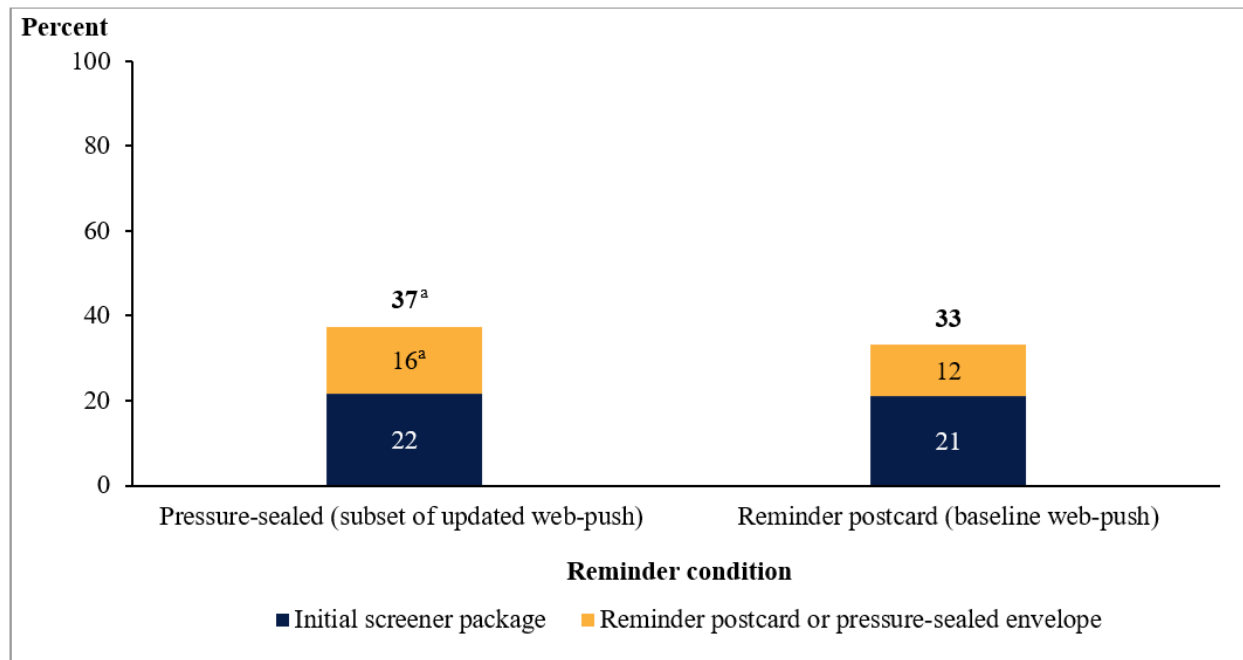
<b>Description</b>	As part of an effort to refine the web-push methodology that was first tested in NHES:2016, the purpose of this analysis was to assess whether a pressure-sealed envelope or a reminder postcard is more effective at encouraging early response in a web-push protocol.
<b>Treatment condition</b>	<b>Subset of updated web-push mailing protocol condition that was sent an advance letter only (n=23,330).</b> These cases received a <b>pressure-sealed envelope</b> after the initial screener package. The interior of the envelope provided the case's web login credentials and survey URL.
<b>Control condition</b>	<b>Baseline web-push (n=40,000).</b> These cases received a <b>reminder postcard</b> after the initial screener package. The postcard mentioned the web response option but did not provide the case's web login credentials and survey URL.
<b>Case selection</b>	Random assignment to conditions.
<b>Results</b>	<ul style="list-style-type: none"> <li>The screener response rate gain after sending a pressure-sealed envelope was 4 percentage points higher than the gain after sending a reminder postcard.</li> <li>There were no subgroups for which sending a reminder postcard was more effective than sending a pressure-sealed envelope.</li> </ul>

NOTE: NHES:2016 is National Household Education Surveys Program of 2016.

### 3.2.1 Screener Response Rate

It was hypothesized that the pressure-sealed envelope would be more effective at increasing response because it included the survey URL and web login credentials. As shown in figure 3.2.1, after the initial screener package, the screener response rate was equivalent in the two conditions (see also table 3.2.1 in appendix E). After the reminder mailing, **the pressure-sealed envelope increased the response rate by about 4 percentage points more than did the reminder postcard.** The bulk of this increase was due to increased response by web.

**Figure 3.2.1. Screener response rate after reminder mailing and percentage point gain in screener response rate after each mailing, by reminder condition: 2019**



<sup>a</sup> Significant difference ( $p < .05$ ) from reminder postcard condition.

NOTE: Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that completed the screener after the specified mailing. Response is attributed to a mailing if the response was received 3 or more days after that mailing was sent and less than 3 days after the next mailing was sent. The pressure-sealed envelope cases are the subset of updated web-push condition that was sent an advance letter only. Unweighted sample size was 21,260 for the reminder postcard condition and 36,700 for the pressure-sealed envelope condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

### 3.2.2 Subgroup Screener Response Rates

Table 3.2.2 in appendix E shows the effect of the pressure-sealed envelope on the screener response rate after the reminder mailing disaggregated by various characteristics of interest. The pressure-sealed envelope led to a statistically significant increase in the screener response rate after that reminder mailing for almost all subgroups. There was not very much differentiation across subgroups in the magnitude of the effect. But there were a few subgroups for which it led to a somewhat larger-than-average increase in response relative to the reminder postcard: (1) addresses with a Black head of household, (2) addresses whose head of household's highest level of education was less than a high school diploma, (3) addresses with four or more adults in the household, and (4) addresses in block groups whose Census Low Response Score was in the third quartile. All of these groups except for households with four or more adults had lower-than-average screener response rates after the reminder mailing; therefore, these results suggest the potential for the pressure-sealed envelope to improve representativeness. There were no subgroups for which the pressure-sealed envelope had a negative impact on the screener response rate.



### 3.3 FedEx Timing Experiment

In previous NHES administrations, reminders sent by FedEx have tended to provide a higher increase in survey response than mailings sent by USPS First-Class Mail. However, it is also more expensive to send FedEx mailings. In addition, the FedEx shipping rate is higher for some addresses than others (e.g., addresses outside of the contiguous United States). The goal of the FedEx timing experiment was to determine the ideal timing for sending FedEx reminders in a web-push design, balancing the potential response gains of sending FedEx mailings earlier with the additional shipping costs associated with doing so.

In NHES:2016, the third screener package was sent via FedEx; this approach was also used for the baseline web-push condition in NHES:2019. Within the updated web-push mailing protocol, sample members were randomly assigned to one of three FedEx timing conditions.

- **Random FedEx second**, in which all cases assigned to this condition were sent the second screener package via FedEx. It was hypothesized that using FedEx for a mailing that offers web response might increase the amount of screener response by web, which is less expensive than paper response and may increase the topical and overall response rates. However, this condition also was expected to increase total FedEx mailing costs because more addresses are sent the second screener package than the third screener package.
- **Random FedEx fourth**, in which all cases assigned to this condition were sent the fourth screener package via FedEx. It was hypothesized that this would reduce total FedEx mailing costs (because fewer addresses are sent the fourth screener package) while still yielding the benefits of including a FedEx mailing, but that it might result in less web screener response compared to the FedEx second condition.
- **Modeled FedEx**, in which a predictive model was used to assign some cases in this condition to receive the second screener package via FedEx and other cases in this condition to receive the fourth screener package via FedEx based on their cost-weighted response propensity score (described in more detail below). It was hypothesized that this condition would offer a compromise between the FedEx second and FedEx fourth conditions and target resources where they were most needed (e.g., for lower response propensity cases). It also was hypothesized that reserving FedEx until the final screener package for some addresses would help offset the cost increase associated with sending the FedEx mailing earlier for the other addresses.

In all conditions, addresses were sent the other reminder mailings by USPS First-Class Mail. Exhibit 3.3 summarizes the design and results of the FedEx timing experiment.

**Exhibit 3.3. FedEx timing experiment key facts**

<b>Description</b>	As part of an effort to refine the web-push methodology that was first tested in NHES:2016, the purpose of this experiment was to determine the ideal timing—in terms of both the response rate and costs—for sending FedEx reminders in a web-push protocol. It also explored whether the ideal timing is different for certain subgroups of the sample, based on their cost-weighted response propensity score (the product of their response propensity and predicted FedEx mailing cost).
<b>Treatment conditions</b>	<p><b>Random FedEx second (n=23,330).</b> Regardless of their cost-weighted response propensity score, all cases assigned to this condition were sent the second screener package by FedEx. This group was part of the updated web-push mailing protocol.</p> <p><b>Random FedEx fourth (n=23,330).</b> Regardless of their cost-weighted response propensity score, all cases assigned to this condition were sent the fourth screener package by FedEx. This group was part of the updated web-push mailing protocol.</p> <p><b>Modeled FedEx (n=23,330).</b> FedEx-high-priority cases assigned to this condition—those with the lowest cost-weighted response propensity scores (n=16,330)—were sent the second screener package by FedEx. FedEx-low-priority cases assigned to this condition—those with higher cost-weighted response propensity scores (n=7,000)—were sent the fourth screener package by FedEx. This group was part of the updated web-push mailing protocol.</p>
<b>Control condition</b>	<b>Baseline web-push (n=40,000).</b> Regardless of their cost-weighted response propensity score, all cases assigned to the baseline web-push condition received a <b>FedEx mailing for the third screener package</b> . Because the NHES mailing protocol typically involves sending the third screener package via FedEx, this was considered the control condition.
<b>Case selection</b>	Random assignment to the four conditions: random FedEx second, random FedEx third (baseline web-push), random FedEx fourth, and modeled FedEx. Within the modeled FedEx condition, model-based assignment to a FedEx second protocol or FedEx fourth protocol.
<b>Results</b>	<ul style="list-style-type: none"> <li>• The cost-weighted response propensity scores were able to identify cases that were FedEx-low-priority and for which delaying FedEx did not reduce their screener response rate.</li> <li>• The screener response rate was similar in the random FedEx second, random FedEx third, and modeled FedEx conditions, but it was 1 to 2 percentage points lower in the random FedEx fourth condition than in the other conditions.</li> <li>• Sending the second package via FedEx in the random FedEx second and modeled FedEx conditions increased early screener response and screener response by web.</li> <li>• The overall response rate was slightly higher in the random FedEx second condition than in the random FedEx third or fourth conditions for both topical surveys (by 2 to 3 percentage points). It also was slightly higher in the modeled FedEx condition than in the random FedEx third or fourth conditions for the ECPP (again, by 2 to 3 percentage points).</li> </ul>

NOTE: NHES:2016 is National Household Education Surveys Program of 2016. ECPP is Early Childhood Program Participation topical survey.

**3.3.1 Classifying FedEx-High-Priority or FedEx-Low-Priority Cases**

Several steps discussed in this section were conducted to classify each sampled address as FedEx-high-priority or FedEx-low-priority.

**3.3.1.1 Developing a cost-weighted response propensity score**

The priority classification for FedEx receipt for each address was based on its cost-weighted response propensity score. This score was produced by multiplying the case's predicted response propensity by its predicted FedEx shipping cost estimate.

- The predicted response propensity for each case was estimated using a binary logistic regression model. The independent variables in this model were address-level variables available on the sampling frame and publicly available estimates of area-level characteristics; the dependent variable was screener response status. Exhibit 3.4 lists the full set of auxiliary variables included in the model. This response propensity model was developed using the NHES:2016 web-push condition as a training dataset and was then applied to the NHES:2019 sample to create a predicted response propensity score for each case in the sample.
- Because FedEx shipping costs were not available for each address in advance of the data collection, a predicted FedEx shipping cost was estimated for each case using a truncated regression model.<sup>3</sup> The independent variables were address-level variables available on the sampling frame that were expected to influence FedEx shipping costs and publicly available estimates of area-level characteristics; the dependent variable was the FedEx shipping cost. Exhibit 3.4 lists the full set of auxiliary variables included in the model. One of the strongest predictors of FedEx shipping cost was the state in which the address was located, with shipping costs being much higher for addresses that were not in the contiguous United States. The truncated regression model was developed using the NHES 2017 web test sample because FedEx cost data were not available for the NHES:2016 sample. It was then applied to the NHES:2019 sample to create a predicted FedEx shipping cost for each case in the NHES:2019 sample.

---

<sup>3</sup> A truncated regression model, a modified version of a linear regression, sets lower and/or upper boundaries for dependent variable prediction. In this case, we used the NHES:2017 web test sample's minimum and maximum FedEx mailing costs as boundaries to avoid getting unrealistically low (e.g., zero or negative) or high predicted costs. P.O. box addresses were randomly assigned to conditions but were excluded from the modeling because FedEx does not deliver to them. We also did not calculate cost-weighted response propensity scores for P.O. box addresses. These cases received regular USPS mailings for both the second and fourth screener packages.

**Exhibit 3.4. Auxiliary variables included in the response propensity and predicted FedEx cost models**

Variable	Source
<b>Response propensity model</b>	
Address flagged as having Hispanic surname	Address-level characteristic from sampling frame
Person age 65 or over present at address	Address-level characteristic from sampling frame
Age of head of household	Address-level characteristic from sampling frame
Education of head of household	Address-level characteristic from sampling frame
Household income	Address-level characteristic from sampling frame
Number of adults in household	Address-level characteristic from sampling frame
Home tenure	Address-level characteristic from sampling frame
Availability of telephone number	Address-level characteristic from sampling frame
Race/ethnicity stratum	Tract-level estimate from ACS
Low Response Score	Block group-level estimate from ACS
Percent Black persons	Block group-level estimate from ACS
Percent persons without high school diploma	Block group-level estimate from ACS
<b>Predicted FedEx shipping cost model</b>	
State	Address-level characteristic from sampling frame
Route type	Address-level characteristic from sampling frame
Dwelling type	Address-level characteristic from sampling frame
Seasonal address indicator	Address-level characteristic from sampling frame
Vacant address indicator	Address-level characteristic from sampling frame
Delivery point type code	Address-level characteristic from sampling frame
Drop point status	Address-level characteristic from sampling frame
Number of addresses associated with drop point	Address-level characteristic from sampling frame
Percent persons that live in a densely populated area	Block group-level estimate from ACS

NOTE: ACS is the U.S. Census Bureau's American Community Survey.

**3.3.1.2 Classifying NHES:2019 cases as FedEx-high-priority or FedEx-low-priority**

After the response propensity score and the predicted FedEx cost for each case had been estimated, each case in the sample was assigned a cost-weighted response propensity score. The sample was sorted by this score in ascending order. Based on the distribution of the scores, there was not a clear cutoff point for differentiating between FedEx-high-priority and FedEx-low-priority cases. Hence, the sample was grouped into deciles, which allowed us to look at the data with more granularity and to meet the operational goal that 75 percent of cases in the modeled FedEx condition that received the second screener package would receive that mailing via FedEx (equivalent to 70 percent of the starting sample in that condition). Cases in lower deciles were expected to have lower response propensities and/or lower FedEx costs, while cases in higher deciles were expected to have higher response propensities and/or higher FedEx costs. Because there was more variability in predicted response propensity than predicted cost, the decile creation was dictated by response propensity more so than cost. For instance, the average predicted response propensity in the highest decile was 2.5 times the predicted response propensity in the lowest decile, while the average FedEx cost estimate in the highest decile was only 1.3 times the average FedEx cost estimate in the lowest decile.

Next, the cases in the three highest deciles (i.e., the top 30 percent of cases with the highest combined response propensities and FedEx costs) were flagged as “FedEx-low-priority,” while the remaining 70 percent of cases with the lowest combined response propensities and FedEx costs were flagged as

“FedEx-high-priority.”<sup>4</sup> All else being equal, the assumption was that the FedEx mailing should be delayed for the cases for which it was predicted to be most costly and/or for which the household had a higher likelihood of responding to the survey without needing a FedEx delivery. Using the cost-weighted response propensity score ensured that the cases predicted to be the most expensive had a higher likelihood of receiving FedEx later, unless they also had a relatively low response propensity score. Appendix D confirms that the cost-weighted response propensity model was successful in identifying FedEx-high-priority and FedEx-low-priority-cases.

### 3.3.1.3 Determining FedEx timing based on cases’ FedEx-priority classification

In the rest of this section, the modeled FedEx condition is compared to the random FedEx second, third, and fourth conditions.

- As shown in exhibit 3.5, in the modeled FedEx condition, high-priority cases were sent the second screener package via FedEx and low-priority cases were sent the fourth screener package via FedEx. Cases that did not receive the package via FedEx for that mailing received it via USPS First-Class Mail.
- In the random FedEx second, third, and fourth conditions, all cases (regardless of FedEx-priority classification) received a FedEx mailing for the second, third, or fourth screener package respectively. When FedEx was not used for a specific mailing, USPS First-Class Mail was used instead. For example, for the second mailing, while cases that were assigned to the FedEx second condition received the mailing via FedEx, those that were in the random FedEx third and fourth conditions received it via USPS First-Class Mail.

**Exhibit 3.5. Timing of FedEx mailing, by experimental condition and FedEx-priority assignment**

FedEx-priority assignment	Experimental condition			
	Random FedEx second	Random FedEx third (control group)	Random FedEx fourth	Modeled FedEx
FedEx-high-priority	Second screener package	Third screener package	Fourth screener package	Second screener package
FedEx-low-priority				Fourth screener package

The analyses in the remainder of this section assess what the impact of FedEx timing is on response outcomes over the four FedEx timing conditions. Is it preferable to send the FedEx mailing earlier or later in the data collection? To address this question, the remainder of this section compares the four FedEx timing conditions to each other.

<sup>4</sup> The distribution of the cost-propensity scores did not show a binary pattern in cases that should be considered high versus low FedEx priority. Hence, we used a conservative cut-off that assigned most of the cases to receive a FedEx second mailing to ensure that the overall NHES:2019 screener response rate was not impacted negatively by delaying the FedEx mailing for a large percentage of the NHES sample; this was achieved by designing the experiment to include a target that at least 75 percent of cases receiving the second screener package would receive it via FedEx.

### 3.3.2 Screener Response Rate

It was hypothesized that sending FedEx mailings earlier in the data collection would result in sample members responding to the screener request more quickly; however, there was not a definitive expectation as to the effect on the final screener response rate. As shown in figure 3.3.1, **the screener response rate was 1 to 2 percentage points lower in the random FedEx fourth condition than in the other conditions. But there was not a statistically significant difference between the screener response rate in the random FedEx second, random FedEx third, or modeled FedEx conditions** (see also table 3.3.1 in appendix E).

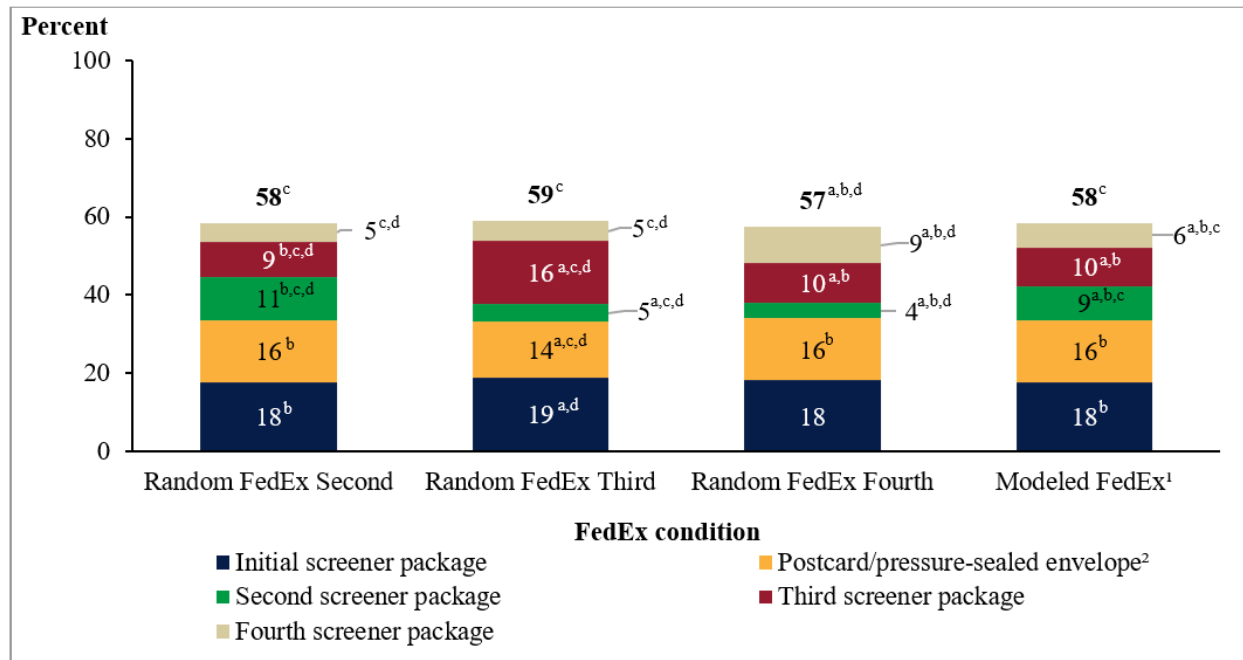
**In all conditions, sending a mailing by FedEx led to a larger increase in screener response after that mailing than sending it by USPS First-Class Mail** (see also table 3.3.2 in appendix E).<sup>5</sup>

- Sending the second screener package via FedEx in the random FedEx second condition increased the response to that mailing by about 6 to 7 percentage points relative to sending it by First-Class mail in the random FedEx third and fourth conditions; **the bulk of this increase was web response.**
- Similarly, sending the third screener package via FedEx in the random FedEx third condition increased the response to that mailing by about 6 to 7 percentage points relative to sending it by First-Class mail in the random FedEx second, random FedEx fourth, and modeled FedEx conditions; however, **the bulk of this increase was paper response.**
- **Sending the fourth screener package by FedEx in the random FedEx fourth condition was slightly less effective.** It increased the response to that mailing by about 4 to 5 percentage points relative to sending it by First-Class mail in the random FedEx second and third conditions; **the bulk of this increase was also paper response.**
- Finally, in the modeled FedEx condition, the increase in response to specific mailings due to FedEx delivery was split between the second and fourth screener packages. Since the use of FedEx in this condition was split between those two screener packages, this result is not unexpected. For example, the increase in response to the second screener package in the modeled FedEx condition (where most cases received FedEx for the second package) compared to the random FedEx third and random FedEx fourth conditions (where all cases received First-Class Mail for the second package) was 4 to 5 percentage points. In addition, the increase in response to the fourth screener package in the modeled FedEx condition (where a small proportion of cases received FedEx for the fourth package) compared to the random FedEx second and third conditions (where all cases received First-Class Mail for the fourth screener package) was about 1 percentage point. The bulk of the increased response to the two mailings was by web.

---

<sup>5</sup> Although all of the other analyses in this report use a 3-day lag period for associating a gain in the response rate with a specific survey mailing, a 1-day lag period was used in the current analysis to account for the faster delivery of FedEx mailings.

**Figure 3.3.1. Final screener response rate and percentage point gain in screener response rate after each mailing, by FedEx condition and mailing: 2019**



<sup>a</sup> Significant difference ( $p < .05$ ) from random FedEx second condition.

<sup>b</sup> Significant difference ( $p < .05$ ) from random FedEx third condition.

<sup>c</sup> Significant difference ( $p < .05$ ) from random FedEx fourth condition.

<sup>d</sup> Significant difference ( $p < .05$ ) from modeled FedEx.

<sup>1</sup> In this condition, a predictive model was used to classify cases into two groups: (1) FedEx-high-priority cases that should receive FedEx earlier in the screener phase (second screener package) and (2) FedEx-low-priority cases that should receive it later in the screener phase (fourth screener package).

<sup>2</sup> Cases in the random FedEx third condition were part of the baseline web-push condition and received a reminder postcard. All other cases were part of the updated mailing protocol condition and received a pressure-sealed envelope.

NOTE: Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that completed the screener after the specified mailing. Response is attributed to a mailing if the response was received 1 or more days after that mailing was sent and less than 1 day after the next mailing was sent. This analysis excludes P.O. box addresses because they are not able to receive FedEx mailings. Unweighted sample size was 21,130 for the random FedEx second condition, 36,450 for the random FedEx third condition, 21,200 for the random FedEx fourth condition, and 21,150 for the modeled FedEx condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

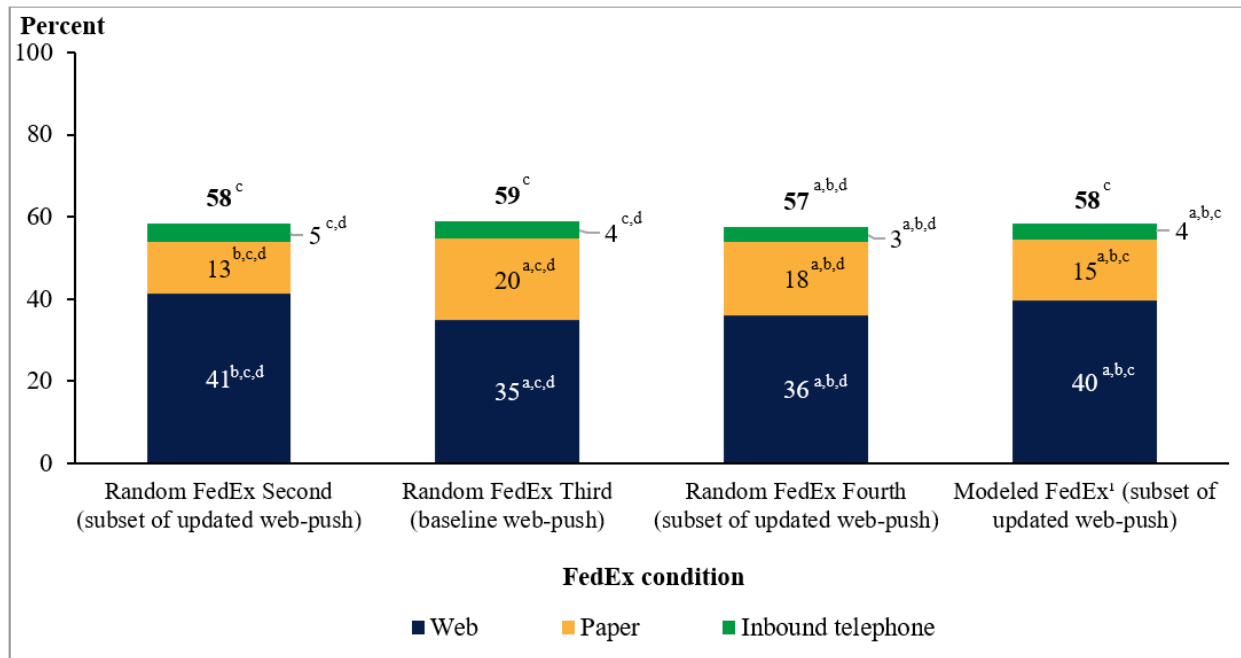
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

### 3.3.3 Screener Mode of Response

It was hypothesized that sending the second screener package via FedEx—prior to providing the paper response option—would increase the amount of screener response received by web. As shown in figure 3.3.2, **sending the second screener package by FedEx increased the web screener response rate by 5 to 6 percentage points, compared to sending the third or fourth screener package by FedEx** (see also table 3.3.1 in appendix E). On the other hand, sending the third or fourth screener package by FedEx increased the paper screener response rate by 5 to 7 percentage points, compared to sending the second screener package by FedEx. As also seen in the prior section, because some cases in this condition received FedEx for the second screener package and others received it for the fourth screener package,

the modeled FedEx condition increase was split between web and paper; it increased the web screener response by 4 to 5 percentage points compared to the random third and fourth FedEx conditions (where all cases received the second screener package by First-Class Mail), and it increased the paper screener response rate by 2 percentage points compared to the random FedEx second condition (where all cases received the second screener package by FedEx).

**Figure 3.3.2. Screener response rate, by FedEx condition and response mode: 2019**



<sup>a</sup> Significant difference ( $p < .05$ ) from random FedEx second condition.

<sup>b</sup> Significant difference ( $p < .05$ ) from random FedEx third condition.

<sup>c</sup> Significant difference ( $p < .05$ ) from random FedEx fourth condition.

<sup>d</sup> Significant difference ( $p < .05$ ) from modeled FedEx condition.

<sup>1</sup> In this condition, a predictive model was used to classify cases into two groups: (1) FedEx-high-priority cases that should receive FedEx earlier in the screener phase (second screener package) and (2) FedEx-low-priority cases that should receive it later in the screener phase (fourth screener package).

NOTE: Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that were respondents to the screener questionnaire. This analysis excludes P.O. box addresses because they are not able to receive FedEx mailings. Unweighted sample size was 21,130 for the random FedEx second condition, 36,450 for the random FedEx third condition, 21,200 for the random FedEx fourth condition, and 21,150 for the modeled FedEx condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

### 3.3.4 Subgroup Screener Response Rates

Table 3.3.3 in appendix E shows the effect of FedEx timing on the screener response rate disaggregated by various characteristics of interest. In general, there was not very much variation in the magnitude of the effect of FedEx timing across subgroups; **even when there was a significant difference in the screener response by condition for specific subgroups, the size of the difference tended to be small (less than 2 percentage points).**



As an exception, sending the second screener package via FedEx was particularly effective at increasing response among the addresses that had the most missing information on the sampling frame (missing for at least four variables); for these addresses, the screener response rate was 4 percentage points higher in the random FedEx second condition than it was in the random FedEx fourth condition. In addition, for addresses that were not in the contiguous United States, sending the second screener package by FedEx increased the screener response compared to using FedEx for the third or fourth screener packages by 9 and 12 percentage points, respectively; however, due to small sample sizes, these differences were not statistically significant. It is also important to note that sending FedEx earlier to these cases may increase costs because FedEx shipping costs are higher for non-contiguous U.S. addresses. However, since only a small number of addresses sampled for the NHES are outside of the contiguous U.S., the magnitude of the impact would be expected to be relatively small.

In contrast, sending the second, rather than the third, screener package via FedEx appeared to reduce the response rate for several subgroups. For example, for paper-sensitive addresses and addresses whose head of household had a graduate degree, the screener response rate was 3 percentage points lower in the random FedEx second condition than it was in the random FedEx third condition. The same pattern was seen for addresses for which the head of household was older than 65. However, since these groups have higher-than-average screener response rates, these reductions seem unlikely to be a threat to representativeness.

### 3.3.5 Topical Eligibility Rate

FedEx timing was not expected to have an effect on the topical eligibility rate. The results show that **FedEx timing did not have an impact on the overall topical eligibility rate or the PFI topical eligibility rate** (see table 3.3.4 in appendix E). Sending the second screener package via FedEx did result in a statistically significant decrease in the ECPP eligibility rate, but the magnitude of this difference was very small (about 1 percentage point).

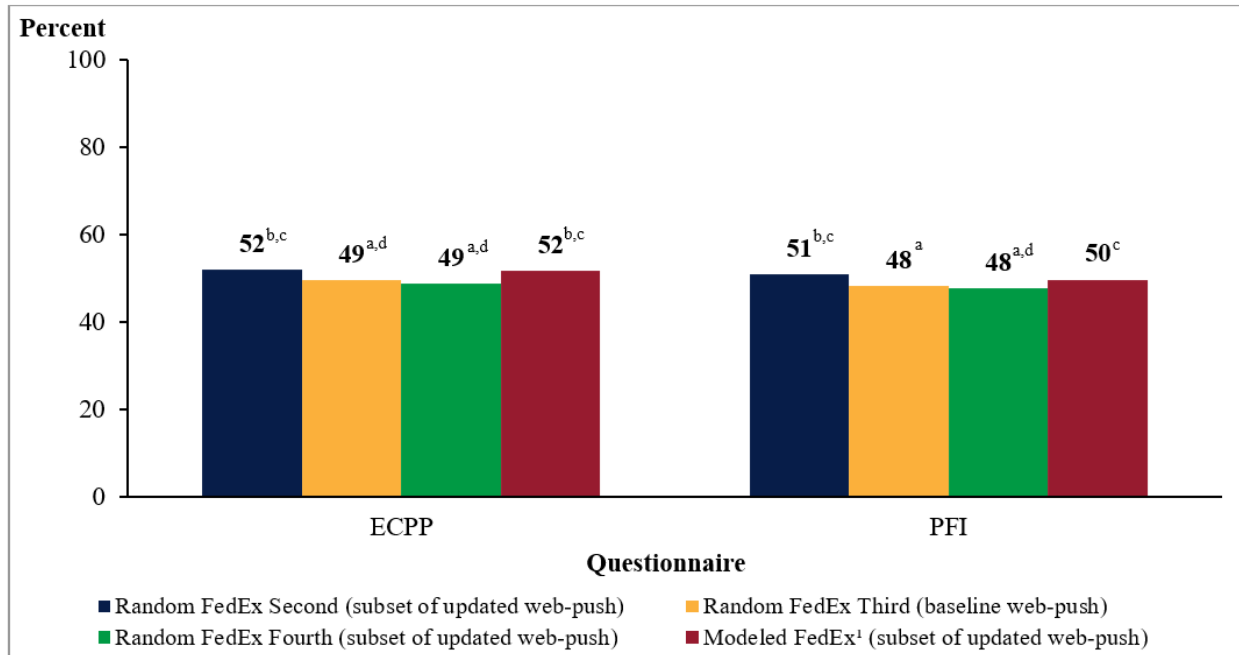
### 3.3.6 Topical Response Rates and Overall Response Rates

FedEx timing was hypothesized to primarily affect screener-phase response. However, to the extent that sending FedEx mailings earlier encouraged earlier response to the screener, and thus increased web response to the screener, it was hypothesized that it might also have a positive downstream effect on the topical and overall response rates. As shown in table 3.3.5 in appendix E, as expected, the ECPP and PFI topical response rates were higher in the random FedEx second and modeled FedEx conditions than in the random FedEx third condition (by about 3 to 5 percentage points). The PFI topical response rate also was higher in the random FedEx second condition than in the random FedEx fourth condition (by about 4 percentage points). There was not a statistically significant difference between the topical response rates in the random FedEx third and fourth conditions.

As shown in figure 3.3.3, **the ECPP overall response rate was higher in the random FedEx second and modeled FedEx conditions than in the random FedEx third and fourth conditions** by 2 to 3 percentage points (see also table 3.3.6 in appendix E). In addition, **the PFI overall response rate was higher in the random FedEx second condition than in the random FedEx third and fourth conditions** (by about 3 percentage points), and the PFI overall response rate was higher in the modeled FedEx condition than in the random FedEx fourth condition (by 2 percentage points). There was not a

statistically significant difference in the overall PFI response rate between the modeled FedEx condition and the random FedEx third condition.

**Figure 3.3.3. Overall response rate, by questionnaire and FedEx condition: 2019**



<sup>a</sup> Significant difference ( $p < .05$ ) from random FedEx second condition.

<sup>b</sup> Significant difference ( $p < .05$ ) from random FedEx third condition.

<sup>c</sup> Significant difference ( $p < .05$ ) from random FedEx fourth condition.

<sup>d</sup> Significant difference ( $p < .05$ ) from modeled FedEx condition.

<sup>1</sup> In this condition, a predictive model was used to classify cases into two groups: (1) FedEx-high-priority cases that should receive FedEx earlier in the screener phase (second screener package) and (2) FedEx-low-priority cases that should receive it later in the screener phase (fourth screener package).

NOTE: ECPP is Early Childhood Program Participation topical survey. PFI is Parent and Family Involvement topical survey. The ECPP and PFI overall response rates are the product of the screener response rates and the respective topical response rate. Screener and topical response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). This analysis excludes P.O. box addresses because they are not able to receive FedEx mailings. Unweighted eligible sample size for the ECPP was 21,130 for the random FedEx second condition, 36,450 for the random FedEx third condition, 21,200 for the random FedEx fourth condition, and 21,150 for the modeled FedEx condition. Unweighted eligible sample size for the PFI was 21,130 for the random FedEx second condition, 36,450 for the random FedEx third condition, 21,200 for the random FedEx fourth condition, and 21,150 for the modeled FedEx condition. Sample sizes have been rounded to the nearest 10.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

Tables 3.3.7 and 3.3.8 in appendix E show the ECPP and PFI overall response rates, respectively, broken out by various subgroups of interest. Differences between the conditions were generally nonsignificant over the conditions as a whole. There was again some evidence that sending the second screener package via FedEx was particularly effective for addresses that were missing information on the sampling frame; for example, it increased the ECPP and PFI overall response rates by 6 to 7 percentage points for addresses that did not have a telephone number on the frame compared to using FedEx for the fourth screener package. There were also a few subgroups for which sending the second mailing via FedEx seemed particularly effective; for example, compared to using FedEx for the third screener package, it increased the ECPP and PFI overall response rates by 5 to 6 percentage points for addresses in tracts with

high poverty rates. However, these comparisons should be interpreted with caution, as the sample for this experiment was not designed for subgroup analyses at the topical level.

### 3.3.7 Cost Per Response

Finally, we assessed the effect of FedEx timing on survey costs. Sending the FedEx mailing earlier—as was done in the random FedEx second condition—could increase costs, as it means more cases receive FedEx mailings; however, it could also reduce costs if it increases the screener response rate by web. In the modeled FedEx condition, delaying FedEx for some cases was expected to help balance any cost increase from sending FedEx earlier for the rest of the cases.

Based on cost estimates produced by the Census Bureau (the NHES data collector) (Varela and Zotti 2020), exhibit 3.6 shows the relative cost per screener response and relative cost per topical response for each pair of FedEx timing conditions.<sup>6</sup> For example, the relative cost per screener response for the random FedEx second condition as compared to the random FedEx third condition was 1.10. This means that the cost per screener response in the random FedEx second condition was 1.10 times the cost per screener response in the random FedEx third condition—that is, the cost per screener response was 10 percent higher in the random FedEx second condition than in the random FedEx third condition. Comparisons were made between each of the conditions as part of assessing the effect of FedEx timing on relative cost per response (that is, no one FedEx timing condition was used as a control).

**Exhibit 3.6. Relative cost per response, by data collection phase and FedEx timing condition: 2019**

FedEx timing condition	Relative cost per screener response	Relative cost per topical response
Random FedEx second versus random FedEx third	1.10	1.01
Random FedEx second versus random FedEx fourth	1.06	1.01
Random FedEx second versus modeled FedEx	1.02	1.02
Modeled FedEx versus random FedEx third	1.07	0.99
Modeled FedEx versus random FedEx fourth	1.04	0.98
Random FedEx fourth versus random FedEx third	1.03	1.01

NOTE: The relative cost is the ratio that is created when comparing the cost per response in the two conditions (e.g., the cost per screener response in random FedEx second condition divided by the cost per screener response in the random FedEx third condition).

SOURCE: Independent tabulations of National Household Education Survey (NHES) 2019 data based on Varela and Zotti (2020).

**Sending the FedEx mailing earlier for all cases.** At the screener phase, sending the second screener package by FedEx for all cases (as was done in the random FedEx second condition) increased the cost per screener response by 6 to 10 percent compared to sending a later mailing by FedEx for all cases (as was done in the random FedEx third and fourth conditions). This is likely driven by the increased number of sampled addresses that incur FedEx shipping costs when the FedEx mailing is sent earlier.

<sup>6</sup> While relative costs per response are used here to assess the effect of FedEx timing on cost, it is important to acknowledge that estimated cost per response for a subset of the NHES:2019 sample may not translate directly to the actual cost of using this approach for the full sample in future administrations.

However, the cost per topical response—which includes both the screener- and topical-phase costs of obtaining a topical response—was similar when the second screener package was sent by FedEx and when a later screener package was sent by FedEx. This difference from the cost per screener response result is likely driven by the increased screener response rate by web when the FedEx mailing is sent earlier, which reduces the need to send paper topical mailings and thus reduces topical-phase costs. As shown in figure 3.3.3 above, the overall ECPP and PFI response rates were higher in the random FedEx second condition than in the random FedEx third and fourth conditions. Sending the FedEx mailing earlier for all cases increased the overall response rate without an increase in the cost per topical response.

**Using a modeled approach.** Compared to the random FedEx second condition, the modeled FedEx condition delayed the FedEx mailing for low-FedEx-priority cases; this was expected to reduce costs (since fewer cases would receive the FedEx mailing). As shown in exhibit 3.6, the cost per screener response and cost per topical response were slightly lower in the modeled FedEx condition than in the random FedEx second condition (by 2 percent). In addition, the screener and overall response rates were equivalent in the modeled FedEx and random FedEx second conditions; this suggests the potential for slight cost savings without a negative impact on response rates. However, sending FedEx mailings at different times for different sample members increases the complexity of the data collection and may lead to increases in other costs for the survey program that are not captured in cost-per-case estimates reported here, such as certain types of labor costs.

## Chapter 4. Attracting Response From Specific Demographic Subgroups

A final set of NHES:2019 experiments tested whether it was possible to increase response among specific demographic subgroups by sending more tailored contact materials.

- **Households without NHES-eligible children:** Although some households without children might assume that an education-related survey is not relevant to them, it is still very important for these households to respond to the screener questionnaire. Can we increase response from such households by giving them an opportunity to “opt out” on the cover page of the paper screener questionnaire?
- **Spanish-speaking households:** Spanish-speaking households tend to have lower NHES response rates than non-Spanish-speaking households. Can we increase response among Spanish-speaking households—and thus reduce nonresponse bias— by providing materials tailored specifically to them?

This chapter presents the results of experiments that tested the use of mailings that were specially tailored to improve screener response rates within these subgroups:

- The **opt-out screener materials experiment**, which aimed to improve screener response rates among households without NHES-eligible children.
- The **targeted Spanish-speaking-household materials experiment**, which aimed to improve screener response rates among Spanish-speaking households.

These experiments varied the wording and design of the screener materials. All experimental conditions discussed in this chapter used a web-push protocol in which the first two screener packages included an invitation to complete the screener via the web, while the final two packages included a paper screener questionnaire.

### 4.1 Opt-Out Screener Materials Experiment

The goal of the opt-out screener materials experiment was to test survey materials that would encourage households without NHES-eligible children to respond to the screener. The experiment was motivated by the hypothesis that households without children might assume that an education-related survey is relevant only to households with children and would thus be less likely to complete the screener. Prior experimental research provides some evidence that sample members are less likely to respond to surveys whose topic is less personally salient to them (Groves, Presser, and Dipko 2004). In keeping with this principle, a National Institute of Statistical Sciences, (NISS) expert panel report recommended that the contact materials and recruitment strategies in NCES surveys be tailored to “communicate effectively with the specific target population(s), based on understanding of what motivates response from that population” (NISS 2018, p. 3).

Although screener respondents without NHES-eligible children are not sampled for a topical survey, their screener completion is important because the overall response rate to the NHES is the product of the

response rates to both the screener and topical phases. Screener response from households without eligible children also enables more precise weighting to be developed for the final survey data, as it reduces the variability of nonresponse-adjusted weights. Furthermore, screener completion allows NCES to verify these sample members' ineligibility for the topical phase and stop sending follow-up screener mailings to households without eligible children.

Exhibit 4.1 summarizes the design and results of the opt-out screener materials experiment.

#### Exhibit 4.1. Opt-out screener materials experiment key facts

<b>Description</b>	The purpose of this experiment was to increase screener response among households without NHES-eligible children. It tested the effectiveness of using materials that emphasize that completing the screener is low burden for households without children.
<b>Treatment condition</b>	<b>Opt-out screener materials (n=10,000).</b> The mailings sent to cases assigned to this condition indicated that the screener is low burden for households without children. The third and fourth screener packages included an "opt-out" version of the paper screener questionnaire, which moved the first question to the cover of the booklet and emphasized that this was the only item necessary for households without children to complete. The contact procedures for this condition were the same as one of the updated web-push mailing protocol conditions: specifically, this condition received an advance letter, a pressure-sealed envelope, and FedEx at the second screener package.
<b>Control condition</b>	<b>Subset of updated web-push mailing protocol condition that received an advance letter only and Fedex at the second screener package (n=7,780).</b> The mailings sent to cases assigned to this condition did not indicate that the screener is low burden for households without children. The third and fourth screener packages included the "regular" paper screener questionnaire in which all items were presented inside the booklet. This control group was one of the updated web-push mailing protocol conditions, receiving an advance letter, a pressure-sealed envelope, and FedEx at the second screener package.
<b>Case selection</b>	Random assignment to conditions.
<b>Results</b>	<ul style="list-style-type: none"> <li>The opt-out screener materials increased the screener response rate among households flagged on the sampling frame as having children, but they did not have an effect among households not flagged as having children or on the screener response rate overall. This was the opposite of the expected results.</li> <li>The opt-out screener materials did not affect the proportion of screener respondents reporting ECPP- or PFI-eligible children.</li> </ul>

NOTE: NHES is National Household Education Surveys Program. ECPP is Early Childhood Program Participation topical survey. PFI is Parent and Family Involvement topical survey.

For the cases randomly assigned to the opt-out screener materials condition, the screener package cover letters included additional language to emphasize that the survey was very short for households without children. For example, the cover letter for the initial screener package stated, "If there are no children in your household, this will take only a minute or two." The cases in the regular screener materials condition received the standard NHES:2019 web-push letters without this extra sentence. The screener contact materials used for both groups are provided in appendix B.

The first two screener packages were web-only survey invitations, so the largest difference in the opt-out screener packages occurred for the third and fourth screener packages, which included a paper questionnaire. For cases assigned to the opt-out screener materials condition, the paper screener questionnaire sent with the third and fourth screener packages was modified to encourage response from households without children. Specifically, the first item asking whether the household had any children was included on the front cover of the screener booklet, allowing households without children to immediately see that they only needed to respond to that item. The control condition used the

standard NHES:2019 paper screener questionnaire with the first item on the inside of the front cover. The paper screeners used for both conditions are provided in appendix C.

The opt-out screener materials were designed to increase response among households without NHES-eligible children. However, whether a household has an eligible child cannot be known with certainty until the household completes the screener. The NHES sample frame does include an indicator for the presence of children (the “child flag”) that is appended by the frame vendor, Marketing Systems Group. However, as discussed further in section 4.1.3, this flag is subject to error: not all child-flagged addresses report NHES-eligible children on the screener, and some non-child-flagged addresses report NHES-eligible children. In particular, the non-child-flagged group includes both addresses that are believed not to have any children present and addresses for which it is unknown whether children are present.

A primary concern about using the opt-out screener materials for all sampled addresses is that the materials could decrease response among households with children by emphasizing that their survey burden would be much higher than it would be for households without children. The opt-out paper screener emphasizes that the survey is just one question for most sampled households. Would households with children, then, feel it unfair to be given a longer survey than households without children? Because of this concern, and the potential for error in the child flag, the opt-out screener materials were tested for a representative sample of all households. This allowed for a more complete evaluation of the impact of the opt-out screener materials on households with and without children, with the understanding that, if the materials were found to have the intended effect (higher response rates among non-child-flagged households and no impact on child-flagged households), they could be targeted to non-child-flagged households in future NHES administrations.

When the condition is considered as a whole, the opt-out screener materials did not significantly change the screener response rate, topical eligibility rates, or topical response rates. However, some differences in effects were observed between child-flagged and non-child-flagged households. Therefore, this section presents results separately for child-flagged and non-child-flagged households to assess the impact of the opt-out screener materials both on the group for which the materials were tailored (non-child-flagged households) and the group for which it was anticipated that the materials might not work as well (child-flagged households). Despite the known error in the child flag, it remains the best available data for examining whether the effect of the opt-out screener materials differs between households with and without children, because the actual presence of NHES-eligible children is known for screener respondents but not for nonrespondents.

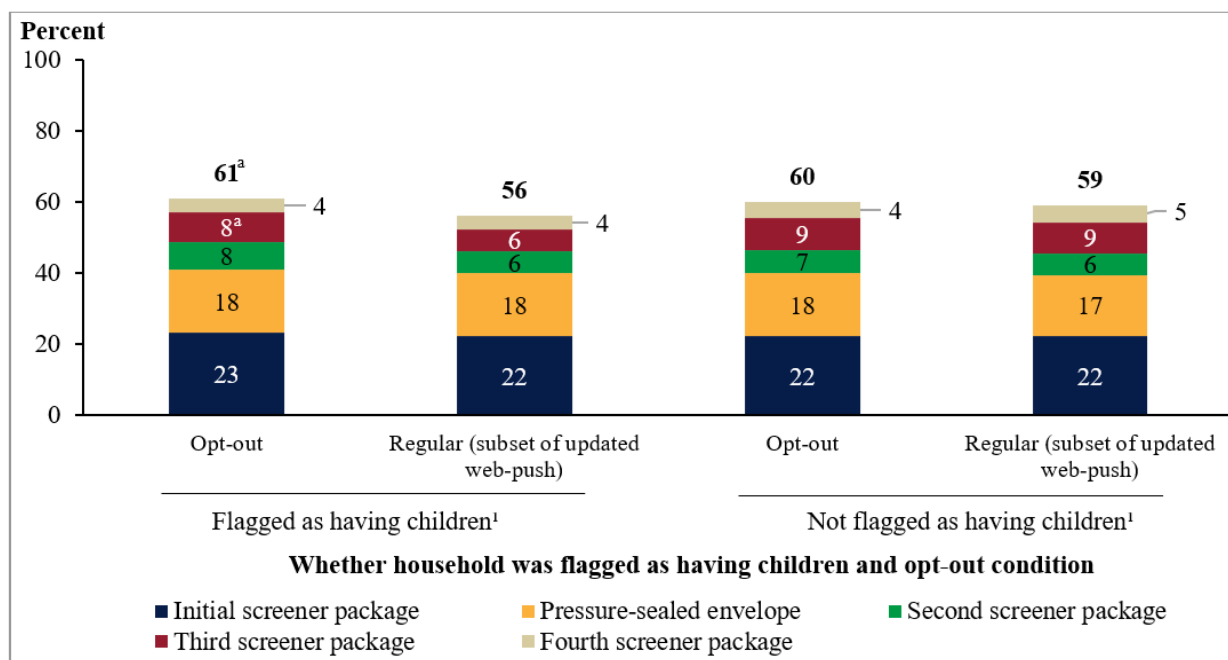
### 4.1.1 Screener Response Rate

It was hypothesized that the opt-out screener materials would increase the screener response rate among households without children while having little or no effect on the screener response rate among households with children.

As shown in figure 4.1.1, the results were the opposite of expectations (see also table 4.1.1 in appendix E). **Among non-child-flagged households, the opt-out screener materials did not have a significant effect** on the screener response rate. **Among child-flagged households, the opt-out screener materials increased the screener response rate by 5 percentage points.** A possible explanation for this unexpected pattern is that the opt-out wording may have primed *all* households, regardless of whether they have children, to expect the survey to be low burden. In this way, the opt-out wording

could have encouraged *greater* response from the group most likely to find the survey salient (households with children), rather than, as originally expected, the group less likely to find the survey salient (households without children). An alternate explanation may be that this opt-out wording made it clearer that the survey primarily focused on households with children, making the screener request even more salient for households with children without any discernable effect on households without children.

**Figure 4.1.1. Final screener response rate and percentage point gain in screener response rate after each mailing, by whether household was flagged as having children, opt-out condition, and mailing: 2019**



<sup>a</sup> Significant difference ( $p < .05$ ) from regular screener materials condition.

<sup>1</sup> The address-based sampling frame used for NHES:2019 included a flag indicating that there were likely to be children living at certain addresses. Addresses lacking this flag are less likely to have children living there. However, it is important to note that the nonflagged group also includes those addresses where the frame vendor could not determine whether children lived there. In addition, prior NHES research has found that there is some error in this flag variable.

NOTE: Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that completed the screener after the specified mailing. Response is attributed to a mailing if the response was received 3 or more days after that mailing was sent and less than 3 days after the next mailing was sent. "Regular" condition cases are the subset of the updated web-push condition that received an advance letter only and Fedex at the second screener package. Unweighted eligible sample size was 1,760 for cases flagged as having children in the opt-out screener materials condition, 7,350 for cases not flagged as having children in the opt-out screener materials condition, 1,350 for cases flagged as having children in the regular screener materials condition, and 5,730 for cases not flagged as having children in the regular screener materials condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

Among child-flagged households, increased response to the third screener package accounted for about half of the total increase in the final response rate (figure 4.1.1; see also table 4.1.2 in appendix E). In fact, this was the only individual mailing that showed a statistically significant increase in the incremental

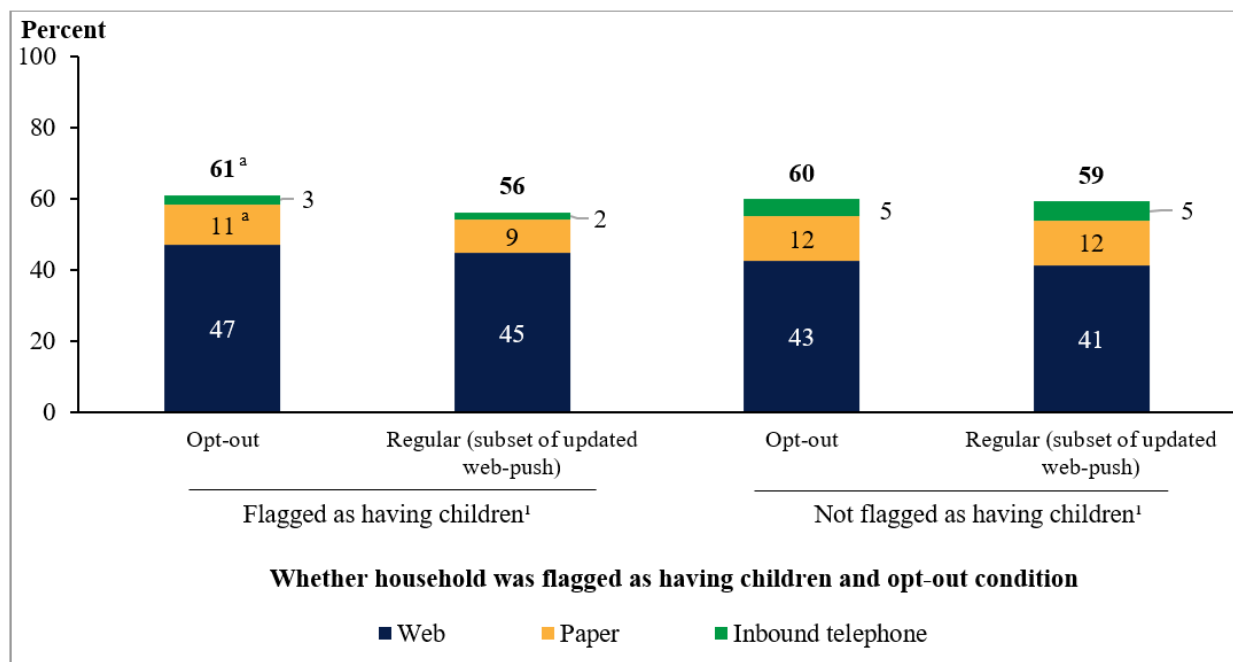


response rate. This may be attributable to the fact that the third screener package was the first package that offered a paper screener questionnaire and that the difference between the opt-out and regular paper questionnaires was more prominent than the differences in the earlier web-only letters. In addition, table 4.1.3 in appendix E shows the effect of the opt-out screener condition on the screener response rate broken out by various subgroups of interest.

### 4.1.2 Screener Mode of Response

Figure 4.1.2 (see also table 4.1.1 in appendix E) disaggregates the screener response rate by the mode of response. Among households flagged as having children, the increase in the response rate was split between web and paper response (though only the paper increase was statistically significant), consistent with the significant increase in response to the third screener package (the first to offer a paper questionnaire). However, for non-child-flagged addresses, there was not a statistically significant change in any of the mode-specific response rates.

**Figure 4.1.2. Screener response rate by whether household was flagged as having children, opt-out condition, and response mode: 2019**



<sup>a</sup> Significant difference ( $p < .05$ ) from regular materials condition.

<sup>1</sup> The address-based sampling frame used for NHES:2019 included a flag indicating that there were likely to be children living at certain addresses. Addresses lacking this flag are less likely to have children living there. However, it is important to note that the nonflagged group also includes those addresses where the frame vendor could not determine whether children lived there. In addition, prior NHES research has found that there is some error in this flag variable.

NOTE: Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that were respondents to the screener questionnaire. "Regular" condition cases are the subset of the updated web-push condition that received an advance letter only and Fedex at the second screener package. Unweighted eligible sample size was 1,760 for cases flagged as having children in the opt-out screener materials condition, 7,350 for cases not flagged as having children in the opt-out screener materials condition, 1,350 for cases flagged as having children in the regular screener materials condition, and 5,730 for cases not flagged as having children in the regular screener materials condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

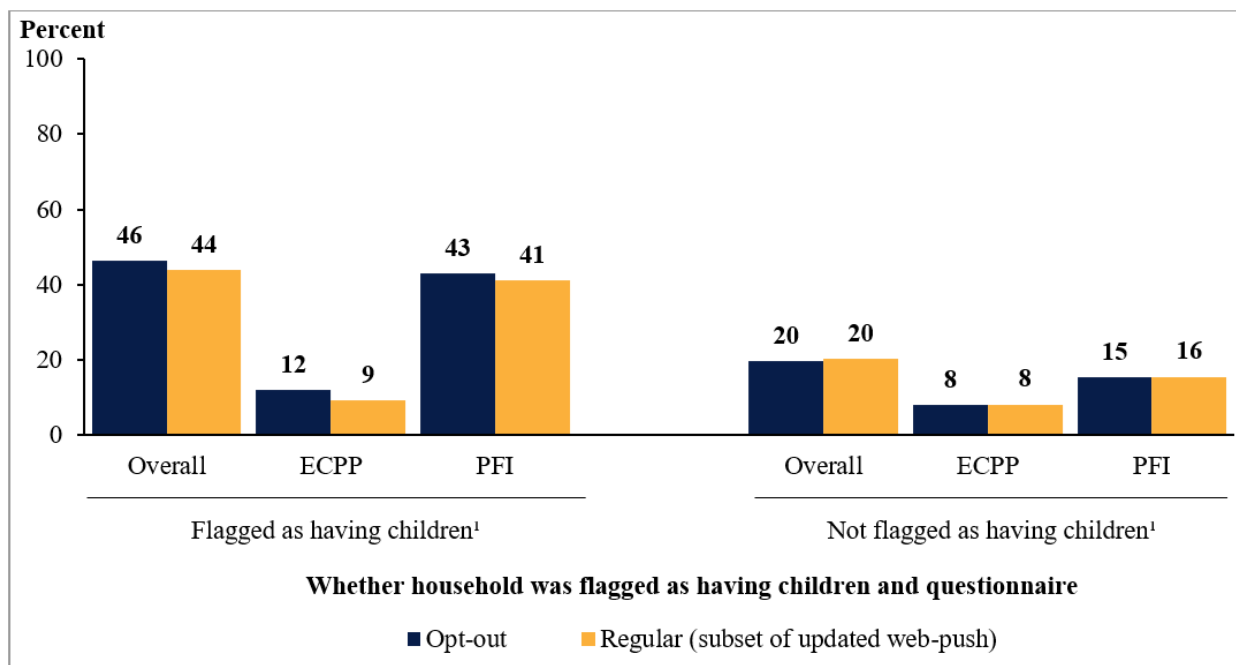
### 4.1.3 Topical Eligibility Rate

Figure 4.1.3 shows the topical eligibility rates for child-flagged and non-child-flagged households, comparing the opt-out screener materials to the regular screener materials (see also table 4.1.4 in appendix E). This comparison is of interest for two reasons.

First, it measures the accuracy of the child flag. As shown in the figure, this flag is subject to substantial measurement error. While the eligibility rate was higher among child-flagged households than among non-child-flagged households, the proportion of child-flagged households that reported NHES-eligible children was under 50 percent in both conditions, implying that the child flag captures a significant number of addresses that do not have NHES-eligible children. Similarly, in both conditions, approximately 20 percent of non-child-flagged households reported NHES-eligible children, implying that the child flag misses a considerable number of addresses that do have eligible children. This may be driven by the fact that non-child-flagged households include both those for which the sample frame vendor believes no children to be present as well as those for which the presence of children cannot be determined. Consequently, if the child flag were used to target opt-out screener materials based on the expected presence of children, there would be substantial inaccuracy in the targeting.

Second, figure 4.1.3 indicates that the opt-out screener materials did not significantly change the topical eligibility rate among either child-flagged or non-child-flagged households. This is relevant because, conceivably, the opt-out materials could have affected households' willingness to accurately report the presence of children. For example, the opt-out screener materials made it clear that the survey would be shorter for households not reporting children, which could have encouraged households with children to erroneously report that they did not have children. Any such "backfire" effect would reduce the topical eligibility rate. In practice, however, the results within the subgroup of child-flagged households suggest little reason for concern over a potential backfire effect. Within this subgroup—which showed a higher screener response rate under the opt-out screener materials condition—the overall topical eligibility rate was not significantly different from the topical eligibility rate in the opt-out screener materials condition. This result suggests that the increase in response among child-flagged households did not come at the expense of reduced reporting of eligible children among these households.

**Figure 4.1.3. Percentage of screener respondents who reported at least one child eligible for the topical surveys, by whether household was flagged as having children, questionnaire, and opt-out condition: 2019**



<sup>1</sup> The address-based sampling frame used for NHES:2019 included a flag indicating that there were likely to be children living at certain addresses. Addresses lacking this flag are less likely to have children living there. However, it is important to note that the nonflagged group also includes those addresses where the frame vendor could not determine whether children lived there. In addition, prior NHES research has found that there is some error in this flag variable.

NOTE: ECPP is Early Childhood Program Participation topical survey. PFI is Parent and Family Involvement topical survey. The overall topical eligibility rate represents the percentage of screener respondent households that reported at least one child on the screener who was eligible for a topical survey. The ECPP and PFI topical eligibility rates each represent the percentage of screener respondents that reported at least one child on the screener who was eligible for that particular topical survey. Screener respondent households may have been eligible for more than one topical survey; as a result, the topical-specific results do not sum to the overall result. "Regular" condition cases are the subset of the updated web-push condition that received an advance letter only and Fedex at the second screener package. Unweighted eligible sample size was 1,050 for cases flagged as having children in the opt-out screener materials condition, 4,270 for cases not flagged as having children in the opt-out screener materials condition, 740 for cases flagged as having children in the regular screener materials condition, and 3,290 for cases not flagged as having children in the regular screener materials condition. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

For reference, tables 4.1.5 and 4.1.6 in appendix E show the topical and overall response rates (respectively), disaggregated by the child flag, for the opt-out and regular screener materials conditions. The topical and overall response rates did not differ between the opt-out and regular screener materials conditions among either child-flagged or non-child-flagged households. Note, however, that the disaggregation by the child flag is of lesser interest for topical-phase results, because all households sampled for a topical survey are households with NHES-eligible children. Thus, non-child-flagged households in the topical sample, by definition, are households for which the child flag was inaccurate.

## 4.2 Targeted Spanish-Speaking-Household Materials Experiment

Historically, Spanish-speaking households have shown lower-than-average response rates to the NHES despite the use of bilingual mailings for likely-Spanish-speaking addresses. For example, in the NHES:2016 web-push condition, addresses that were designated to receive at least one bilingual mailing had a screener response rate of 55 percent, compared to 63 percent for addresses that were not designated to receive any bilingual mailings.

The targeted Spanish-speaking-household materials experiment sought to determine whether further tailoring of the wording and design of bilingual mailing materials could improve response rates from this critical subgroup. If successful, this tailoring would have the potential to improve representativeness by reducing the difference in response rates between likely-Spanish-speaking and not-likely-Spanish-speaking households.

The tailored materials used in this experiment were informed by seven in-person focus groups conducted with Spanish-speaking persons in Rockville, Maryland, and Denver, Colorado, from April through June of 2018. In a first set of focus groups, participants described why they do or do not participate in surveys and discussed education-related themes of importance to them. In a second set of focus groups, participants were provided with drafts of various potential designs for the NHES screener materials. The participants in these focus groups responded positively to the design features that were ultimately incorporated into the tailored materials, including endorsements from Hispanic-affiliated organizations and graphic icons emphasizing that the survey is voluntary and trusted. Additional detail about these design features is provided below. Exhibit 4.2 summarizes the design and results of this experiment.

**Exhibit 4.2. Targeted Spanish-speaking-household materials experiment key facts**

<b>Description</b>	In an effort to increase response among Spanish-speaking households, the purpose of this experiment was to assess the effectiveness of sending tailored materials to households that, based on information available on the sampling frame, were likely to be Spanish-speaking. These materials were developed to appeal to Spanish-speaking households based on focus groups conducted with Spanish speakers. Only the contact materials were changed for the experiment, not the survey response modes or the sequence in which they were offered.
<b>Treatment condition</b>	<b>Targeted likely-Spanish-speaking-household materials (n=15,000).</b> The likely-Spanish-speaking cases in this condition (n=3,370, identified using variables available on the sampling frame) received tailored materials. The not-likely-Spanish-speaking cases in this condition (n=11,630) received the same materials as were used in the baseline web-push condition. All cases assigned to this condition were part of a web-push protocol in which the first two screener packages requested web response and the last two screener packages included a paper questionnaire.
<b>Control condition</b>	<b>Baseline web-push (n=40,000).</b> Regardless of whether they were likely to be Spanish-speaking, none of the cases received the tailored materials. All cases assigned to this condition were part of a web-push protocol in which the first two screener packages requested web response and the last two screener packages offered a paper questionnaire.
<b>Case selection</b>	Random assignment to conditions. Within the targeted likely-Spanish-speaking-household condition, assignment to tailored or baseline materials was based on variables available on the sampling frame.
<b>Results</b>	<ul style="list-style-type: none"> <li>The tailored materials reduced the screener and overall response rates among likely-Spanish-speaking households, the opposite of the expected effect.</li> <li>Among likely-Spanish-speaking households, there were no subgroups within which the tailored materials had a positive effect on the screener response rate.</li> <li>The tailored materials did not increase the percentage of respondents who responded to the survey in Spanish.</li> </ul>

**Criteria used to flag likely-Spanish-speaking households.** A total of 15,000 addresses were randomly assigned to the targeted likely-Spanish-speaking-household materials condition (subsequently referred to as “the targeted materials condition”). Within this condition, likely-Spanish-speaking households were sent the tailored materials with the screener and topical mailings. Not-likely-Spanish-speaking households within the targeted materials condition were sent the same screener and topical materials as in the baseline web-push condition. Thus, the treatment applied only to likely-Spanish-speaking households.

The language spoken by the residents of an address is not directly observable prior to data collection. Therefore, “likely-Spanish-speaking households” needed to be identified using proxy information appended by the sample frame vendor or from publicly available Census data. For this experiment, likely-Spanish-speaking households were defined as addresses that met at least one of the following criteria:

- The address was flagged by the sample frame vendor as containing one or more persons with a Hispanic surname.<sup>1</sup>
- The address was in a Census tract with 40 percent or more Hispanic persons.
- The address was in a Census tract in which 10 percent or more of households have Spanish as their primary language and “limited English-speaking” status.<sup>2</sup>

Approximately 22 percent of NHES:2019 sampled addresses were classified as likely-Spanish-speaking households based on these criteria. Due to the reliance on address- and neighborhood-level proxy data, it was expected that not all such addresses would, in fact, have Spanish-speaking residents and that some Spanish-speaking households would not be captured by these criteria. The criteria attempted to balance the twin goals of capturing as many Spanish speakers as possible while minimizing the number of non-Spanish speakers who would receive materials tailored to Spanish speakers.

Separate from this experiment, NHES operations identify all sampled addresses as either an address that will receive an English-language screener package or an address that will receive a bilingual screener package. These criteria evolve throughout the screener data collection such that the smallest number of households are flagged to receive bilingual packages at the initial screener package, and the pool of potential bilingual screener package recipients expands as nonresponse follow-up packages are mailed. Because of the potential for the results of this experiment to be confounded by whether an address received a bilingual package, analysis of the experimental results utilizes a smaller analytic sample than the sample of all cases receiving materials tailored for likely-Spanish-speaking households. Specifically, the Census tract criterion was made narrower, such that the address had to be in a Census tract that both

---

<sup>1</sup> The frame vendor uses a proprietary method to append a flag to the frame that identifies households that include persons with a Hispanic surname (based on two undisclosed commercial data sources). The flag indicates whether any person in the household (as opposed to just the head of the household) has a Hispanic surname.

<sup>2</sup> The Census Bureau defines a “limited English-speaking” household as one in which no member age 14 or older either (1) speaks only English or (2) speaks a non-English language but also speaks English “very well.”

had 40 percent or more Hispanic persons *and* did not have 25 percent or more Black persons.<sup>3</sup> This ensured that *all* likely-Spanish-speaking cases in both conditions included in this analysis received bilingual versions of *all* screener mailings. As a result of this restriction, the results presented in this chapter exclude approximately 120 cases from the baseline web-push condition and 40 from the targeted materials condition.

**Targeted materials protocol.** In the targeted materials condition, likely-Spanish-speaking households received bilingual versions of all screener contacts. The bilingual mailings included both English and Spanish versions of the letters and questionnaires. They included several features that were expected to improve response rates among likely-Spanish-speaking households (see appendix B for the contact materials and appendix C for the paper questionnaire). These features included:

- Icons on the envelopes emphasizing that the survey is voluntary, is about the respondent’s community, and can be trusted (see exhibit 4.3). Additionally, the text “Please respond within two weeks” was shown in both English and Spanish on the envelopes, with the Spanish text coming first. All screener packages were sent in a full-size (9x13-inch) envelope.
- Spanish-first presentation of the bilingual letters and questionnaires, such that the household would see the Spanish version first after opening the envelope and would need to flip the letter over to see the English text.
- A bilingual addressee line on the letters that showed through the envelope’s address window.
- In the initial and third screener mailing packages, customized letters containing the same icons used on the envelopes, a list of Hispanic-affiliated nongovernmental organizations that had endorsed the survey, and language emphasizing that responses are important.
- In the auto-call reminder that occurred concurrent with the third screener package, the message began with “Hola!” instead of “Hello!”, though the remainder of the message was in English.
- In the third and fourth screener packages, an alternate version of the paper screener questionnaire with images of Hispanic families on the cover, instead of the images of families from a wide range of racial/ethnic backgrounds that were used in the baseline version of the questionnaire.

---

<sup>3</sup> Tracts with 40 percent or more Hispanic persons and not 25 percent or more Black persons correspond to the Hispanic sampling stratum used for the screener-level sample design in the NHES. Under the baseline web-push condition, bilingual materials were used for cases that are in the Hispanic sampling stratum; flagged by the vendor as having a Hispanic surname; and/or in Census tracts with 10 percent or more Spanish-speaking, “limited English-speaking” households.

**Exhibit 4.3. Icons used on materials sent to likely-Spanish-speaking households in targeted materials condition**



In the baseline condition, likely-Spanish-speaking households received bilingual mailings that included both English and Spanish versions of the letters and questionnaires. The English letters and questionnaires were the same versions used for not-likely-Spanish-speaking households. The screener packages were arranged such that the household would see the English version of each material before the Spanish version, and the envelopes omitted the icons and bilingual language.

At the topical phase, the mailings for likely-Spanish-speaking households were mostly identical between the two conditions. There were only two differences in the targeted materials condition: (1) the envelopes used to send paper questionnaires included the icons reiterating the messages noted above and (2) the text on the envelopes (“Thank you for your participation. The Census Bureau asks you to complete this final survey.”) was shown in both English and Spanish, with Spanish first.

Exhibit 4.4 summarizes the mailing materials that were used for likely-Spanish-speaking and not-likely-Spanish-speaking households (as defined for the purpose of the analyses discussed in this section), comparing the targeted materials condition to the baseline web-push condition.

**Exhibit 4.4. Assigned mailing materials, by experimental condition and likely-Spanish-speaking classification**

Likely-Spanish-speaking classification <sup>1</sup>	Selected characteristics of materials	Experimental condition	
		Targeted materials condition	Baseline web-push condition
Likely-Spanish-speaking	Language of screener materials	Bilingual for all mailings	Bilingual for all mailings
	Type of screener materials	Tailored	Non-tailored
	Language of topical materials	Spanish if responded to screener in Spanish; otherwise, English	Spanish if responded to screener in Spanish; otherwise, English
	Type of topical envelope for mailings with paper questionnaire	Tailored	Non-tailored
Not-likely-Spanish-speaking	Language of screener materials <sup>2</sup>	Initial screener package: English Second screener package: <ul style="list-style-type: none"> <li>• Bilingual for 12 percent of cases</li> <li>• English for remainder</li> </ul> Third and fourth screener packages: <ul style="list-style-type: none"> <li>• Bilingual for 66 percent of cases</li> <li>• English for remainder</li> </ul>	
	Type of screener materials	Non-tailored	
	Language of topical materials	Spanish if responded to screener in Spanish; otherwise, English	
	Type of topical envelope for mailings with paper questionnaire	Non-tailored	

<sup>1</sup> This exhibit refers to likely-Spanish-speaking households as they were defined for analytic purposes.

<sup>2</sup> Among households not classified as likely-Spanish-speaking, bilingual mailings were used for the second through fourth screener packages if the household was located in a Census tract in which 3 to 10 percent of households had Spanish as their primary language and “limited English speaking” status; and for the third and fourth screener packages if the household was located in a Census tract in which 2 to 3 percent of households had Spanish as their primary language and “limited English-speaking” status, or a Census tract in which at least 2 percent of persons spoke Spanish at home.

This section discusses the results among likely-Spanish-speaking households, to which the experiment applied. All figures show results for likely-Spanish-speaking households, broken out by the assigned condition (targeted materials condition or baseline web-push condition). Regardless of experimental condition, all likely-Spanish-speaking households included in the analyses received bilingual materials for all mailings; hence, the results comparing outcomes by condition for likely-Spanish-speaking households control for the language of the mailings and, more specifically, show the impact of tailoring on survey outcomes. All figures also include results for not-likely-Spanish-speaking households. The results for not-likely-Spanish-speaking households pool both conditions because the experiment did not apply to these households (i.e., as shown in exhibit 4.4, the mailing protocols and materials were the same in both conditions). These results are included for comparison purposes, to indicate whether the tailored materials helped to equalize response outcomes between likely-Spanish-speaking and not-likely-Spanish-speaking households.

Though the tailored materials had large and statistically significant effects within the subgroup to which the experiment applied—likely-Spanish-speaking households—this subgroup constitutes a relatively small share of the full targeted materials condition (approximately 22 percent). Therefore, effects on response



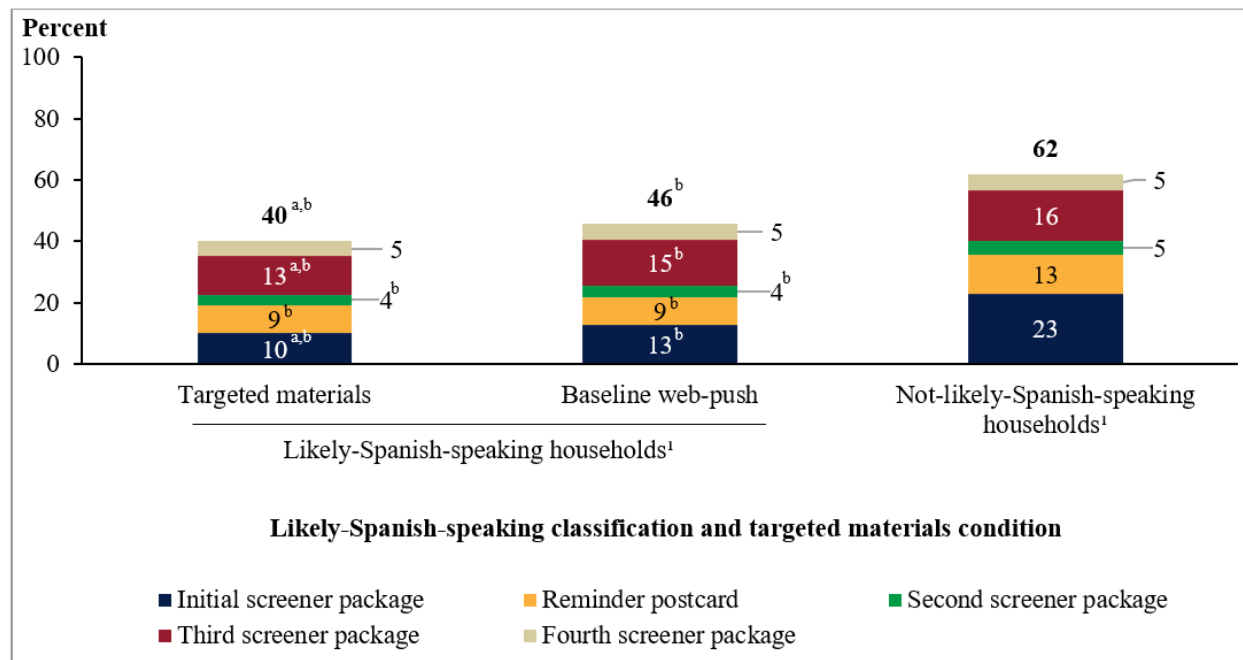
outcomes over the conditions as a whole were much smaller and, in most cases, not statistically significant. For this reason, this section focuses on the results among likely-Spanish-speaking households and does not discuss results over the conditions as a whole.

### 4.2.1 Screener Response Rate

In the baseline web-push condition, the screener response rate was lower among likely-Spanish-speaking households than among not-likely-Spanish-speaking households, as in NHES:2016. It was hypothesized that the tailored materials would increase the screener response rate among likely-Spanish-speaking households and thereby reduce the difference in the screener response rate between likely-Spanish-speaking and not-likely-Spanish-speaking households.

As shown in figure 4.2.1, the actual effect was the opposite of expectations (see also table 4.2.1 in appendix E). **The tailored materials further decreased the screener response rate among likely-Spanish-speaking households, by 6 percentage points.** Consistent with patterns shown in previous NHES administrations, likely-Spanish-speaking households in the baseline web-push condition had a lower screener response rate than not-likely-Spanish-speaking households. Therefore, the tailored materials *increased* the difference in the screener response rate between likely-Spanish-speaking and not-likely-Spanish-speaking households.

**Figure 4.2.1. Final screener response rate and percentage point gain in screener response rate after each mailing, by likely-Spanish-speaking classification, targeted materials condition, and mailing: 2019**



<sup>a</sup> Significant difference ( $p < .05$ ) from baseline web-push condition.

<sup>b</sup> Significant difference ( $p < .05$ ) from not-likely-Spanish-speaking households.

<sup>1</sup> Likely-Spanish-speaking households were identified based on information available on the sampling frame: (1) the address is flagged as having a Hispanic surname, (2) the address is in a tract with 40 percent or more Hispanic persons, or (3) the address is in a tract where 10 percent or more of households have Spanish as their primary language and “limited English-speaking” status.

NOTE: Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RRI). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that completed the screener after the specified mailing. Response is attributed to a mailing if the response was received 3 or more days after that mailing was sent and less than 3 days after the next mailing was sent. Unweighted sample size was 2,890 for likely-Spanish-speaking cases in the targeted materials condition and 8,210 for likely-Spanish-speaking cases in the baseline web-push condition. Unweighted sample size was 38,940 for not-likely-Spanish-speaking cases in the targeted materials and baseline web-push conditions. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

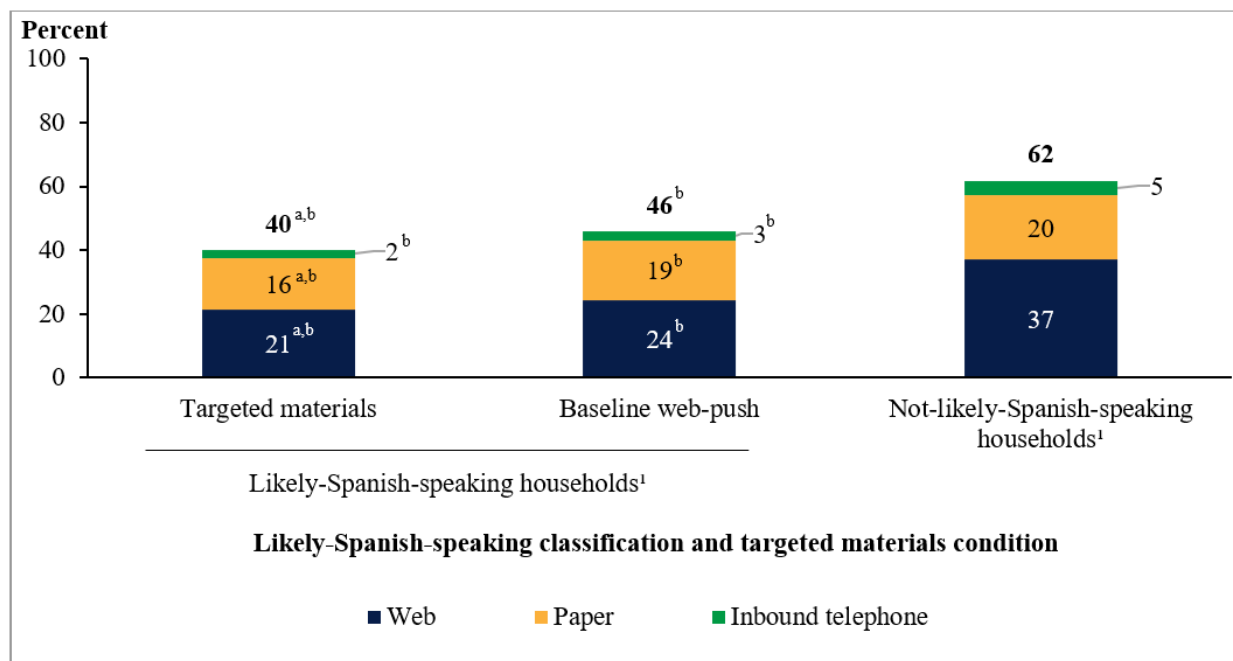
Among likely-Spanish-speaking households, **the initial and third screener packages accounted for the bulk of the decrease in the screener response rate** (see also table 4.2.2 in appendix E). At these mailings, the tailored materials reduced the incremental response rate by about 3 percentage points and 2 percentage points, respectively. These were the only mailings for which the incremental response rate differed significantly between the conditions. This may be because the first and third screener packages were the only packages in which the text of the letters differed between the conditions. Furthermore, these were the first packages offering the web and paper modes, respectively. Therefore, they were the first mailings at which a household would have seen the tailored materials specific to each mode.

At each individual mailing except the fourth screener package, the incremental response rate was significantly lower for likely-Spanish-speaking households (under both conditions) than for not-likely-Spanish-speaking households. This gap was particularly meaningful for the initial screener package, at which the incremental response rate for likely-Spanish-speaking households was more than 10 percentage points lower (under both conditions) than for not-likely-Spanish-speaking households.

### 4.2.2 Screener Mode of Response

As shown in figure 4.2.2, the decrease in the screener response rate under the targeted materials condition was approximately evenly split between web and paper response, with the web screener response rate lower by about 3 percentage points and the paper screener response rate lower by about 2 percentage points among likely-Spanish-speaking households (see also table 4.2.1 in appendix E). This is consistent with the pattern of response rates by individual mailings, as most responses to the initial screener package were by web while most responses to the third screener package were by paper.

While all mode-specific response rates were lower for likely-Spanish-speaking households (under both conditions) than for not-likely-Spanish-speaking households, the gap was particularly large for the response rate by web.

**Figure 4.2.2. Screener response rate, by likely-Spanish-speaking classification, targeted materials condition, and response mode: 2019**

<sup>a</sup> Significant difference ( $p < .05$ ) from baseline web-push condition.

<sup>b</sup> Significant difference ( $p < .05$ ) from not-likely-Spanish-speaking households.

<sup>1</sup> Likely-Spanish-speaking households were identified based on information available on the sampling frame: (1) the address is flagged as having a Hispanic surname, (2) the address is in a tract with 40 percent or more Hispanic persons, or (3) the address is in a tract where 10 percent or more of households have Spanish as their primary language and “limited English-speaking” status.

NOTE: Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RRI). Percentages represent the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that were respondents to the screener questionnaire. Unweighted sample size was 2,890 for likely-Spanish-speaking cases in the targeted condition and 8,210 for likely-Spanish-speaking cases in the baseline web-push condition. Unweighted sample size was 38,940 for not-likely-Spanish-speaking cases in the targeted materials and baseline web-push conditions. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

### 4.2.3 Subgroup Screener Response Rates

Table 4.2.3 in appendix E shows the effect of the targeted materials condition on the screener response rate broken out by various subgroups of interest.

Among likely-Spanish-speaking households, **the targeted materials condition resulted in a lower screener response rate than the baseline web-push condition within all subgroups shown in the table.** That is, there was no subset of likely-Spanish-speaking households for which the tailored materials had a positive effect. However, patterns of variation in the *magnitude* of the decrease point to some hypotheses about the reasons for the unexpected results of this experiment.

First, among likely-Spanish-speaking households, **the response rate decrease was generally smaller than average among addresses whose address-level characteristics suggest a higher probability that the household living at the address was, in fact, Spanish-speaking.** For example, addresses

with a Hispanic head of household (based on vendor frame information) showed a 4-percentage-point decrease under the targeted materials condition, compared to 10 percentage points for addresses with a White head of household. As noted at the beginning of this section, the criteria used to identify likely-Spanish-speaking households included two neighborhood-level characteristics (percent Hispanic and percent Spanish-speaking with limited English-speaking status), as well as one address-level characteristic (the Hispanic surname flag). The neighborhood-level characteristics are inherently imprecise indicators because any specific household living in a heavily Spanish-speaking area may not itself be Spanish-speaking. As discussed in section 4.2.6 below, the tailored materials could plausibly have backfired for households that received them but were not, in fact, Spanish-speaking. This hypothesis is consistent with the fact that the response rate decrease was smaller for subgroups for which *address-level* information suggests a higher likelihood of being Spanish-speaking. At the same time, the fact that the effect was still negative for these subgroups also suggests that the materials themselves did not achieve the intended result even for households more likely to have been accurately targeted.

Second, the **response rate decrease was generally larger than average among addresses whose characteristics were consistent with higher socioeconomic status**. For example, the following patterns were observed:

- **Marital status of head of household:** Addresses with a married head of household showed a decrease of over 8 percentage points, compared to approximately 4 percentage points for addresses with an unmarried head of household.
- **Education of head of household:** Addresses whose head of household had a graduate degree as their highest level of education showed a decrease of over 13 percentage points, much larger than the decreases observed for the other educational attainment categories.
- **Home tenure:** Homeowners showed a decrease of over 7 percentage points, compared to approximately 4 percentage points among renters.
- **Number of adults in the household:** Two-adult households showed a decrease of over 9 percentage points, compared to under 5 percentage points for single-adult households.

### 4.2.4 Topical Eligibility Rate

As shown in table 4.2.4 in appendix E, among likely-Spanish-speaking households, the targeted materials condition did not significantly change the topical eligibility rate. Under both conditions, likely-Spanish-speaking households were much more likely than not-likely-Spanish-speaking households (by about 10 percentage points) to report ECPP- or PFI-eligible children.

### 4.2.5 Topical Response Rates and Overall Response Rates

At the topical phase, the envelopes used for likely-Spanish-speaking households in the targeted materials condition included bilingual text and icons similar to those used at the screener phase. This was the only difference between the targeted materials and baseline web-push conditions at the topical phase.

As shown in table 4.2.5 in appendix E, for both the ECPP and the PFI, the **topical response rate among likely-Spanish-speaking households did not differ significantly between the targeted materials and baseline web-push conditions**. However, the direction of the difference (approximately

4 percentage points lower under the targeted materials condition) was consistent with the direction of the screener-level effect.

Accordingly, **among likely-Spanish-speaking households, the targeted materials decreased the overall response rate by over 6 percentage points** for both the ECPP and PFI (see table 4.2.6 in appendix E). Therefore, the difference in overall response rates between likely-Spanish-speaking and not-likely-Spanish-speaking households became larger.

For reference, tables 4.2.7 and 4.2.8 in appendix E show the ECPP and PFI overall response rates, respectively, broken out by various subgroups of interest. Among likely-Spanish-speaking households, the overall response rates to both topical surveys were lower under the targeted materials condition within nearly all subgroups examined (although statistical significance could often not be confirmed due to small sample sizes). Within nearly all subgroups examined, under both conditions, the overall response rates to both surveys were significantly lower for likely-Spanish speaking households than for not-likely-Spanish-speaking households. Again, however, the sample for this experiment was not designed for subgroup analyses at the topical level, so these results should be interpreted with caution.

#### 4.2.6 Possible Drivers of Results

Overall, therefore, the tailored materials had the opposite of the intended effect, reducing response rates among likely-Spanish-speaking households. This result could be driven by either, or both, of the factors described below.

**Challenges of targeting using address-based sampling (ABS) frames.** Because the NHES screener sample is selected from an ABS frame, likely-Spanish-speaking households needed to be identified using proxy information available on or linkable to the sampling frame. Address-level indicators appended to ABS frames are known to be subject to some inaccuracy (Roth, Han, and Montaquila 2013; West et al. 2015), and neighborhood-level information—like the U.S. Census Bureau’s American Community Survey (ACS) variables used for targeting in this experiment—is an inherently imprecise indicator of household-level characteristics. It is possible that the tailored materials backfired among households that were flagged as likely-Spanish-speaking—based on the limited information available on the sampling frame—but were not, in fact, Hispanic or Spanish-speaking. For example, such households may have assumed from the Spanish-first presentation of the materials that the survey was directed only at Spanish-speaking populations, and thus discarded the mailings without noticing the English-language versions.

**Ineffectiveness of the materials themselves.** Although the tailored materials were created based on focus-group feedback from Spanish speakers, it is possible that, even when households were correctly flagged as Spanish speakers, the tailored materials inadvertently discouraged rather than encouraged response. Notably, the NHES:2019 collection coincided with significant media coverage of the proposed addition of a citizenship question to the Decennial Census. Focus group research by the Census Bureau found evidence that the controversy over the citizenship question substantially increased the reluctance to participate in the Decennial Census among Hispanic respondents, including native-born and naturalized citizens, due to fears about the intended use of the data (Evans et al. 2019). This mistrust may have carried over to the NHES, which uses Census-branded contact materials. In this context, it is possible that the tailored materials backfired by suggesting that the government was aware of individual sample members’ ethnic heritage or language status and was targeting Spanish-speaking households to collect data that could be used against them. In addition, the tailored materials highlighted the voluntary

nature of the survey more than the materials used in the baseline condition; it is possible that this made some households less likely to participate.

Because a household's actual Spanish-speaking status is not observable prior to response, neither of these factors can be confirmed or ruled out with certainty. However, two proxy measures, described below, can provide further insight into the accuracy with which households were designated as likely-Spanish-speaking and whether the tailored materials affected response rates among *actual* Spanish-speaking households.

- **Observed Spanish-speaking status** (i.e., the percentage of screener respondents who completed the screener in Spanish). This is an imperfect indicator of Spanish-speaking status because some Spanish speakers who also speak some English may choose to complete the screener in English. For example, NHES qualitative research (Office of Management and Budget 2018; Medway et al. 2022) finds that some Spanish speakers, based on prior experience with other translated materials, expect Spanish translations to be poorly done. This could condition them to respond in English even when a Spanish questionnaire is available. With this caveat, this rate can be interpreted as a rough proxy for the success of the tailored screener mailings at obtaining responses from households that prefer to respond in Spanish.
- **Self-reported Spanish-speaking status** (i.e., the percentage of topical respondents who self-reported that the sampled child or any adult in the household speaks Spanish at home).<sup>4</sup>

As shown in figure 4.2.3 (results not shown in tables), an examination of these metrics indicates the following:

- In the baseline web-push condition, 9 percent of likely-Spanish-speaking screener respondents completed the screener in Spanish, and 55 percent of likely-Spanish-speaking topical respondents self-reported that Spanish was spoken in the home. The corresponding measures among not-likely-Spanish-speaking cases were under 1 percent and 6 percent, respectively. Therefore, while both measures were (as expected) much higher among likely-Spanish-speaking cases, **there remained substantial imprecision in the targeting**: over 90 percent of likely-Spanish-speaking respondents completed the screener in English, and nearly 45 percent did not self-report that any Spanish was spoken at home.
- Among likely-Spanish-speaking households, the percentage responding to the screener in Spanish was the same in the targeted materials condition as in the baseline web-push condition (9 percent). Similarly, the percentage of likely-Spanish-speaking households that self-reported speaking some Spanish at home did not differ significantly between the conditions. Thus, **there**

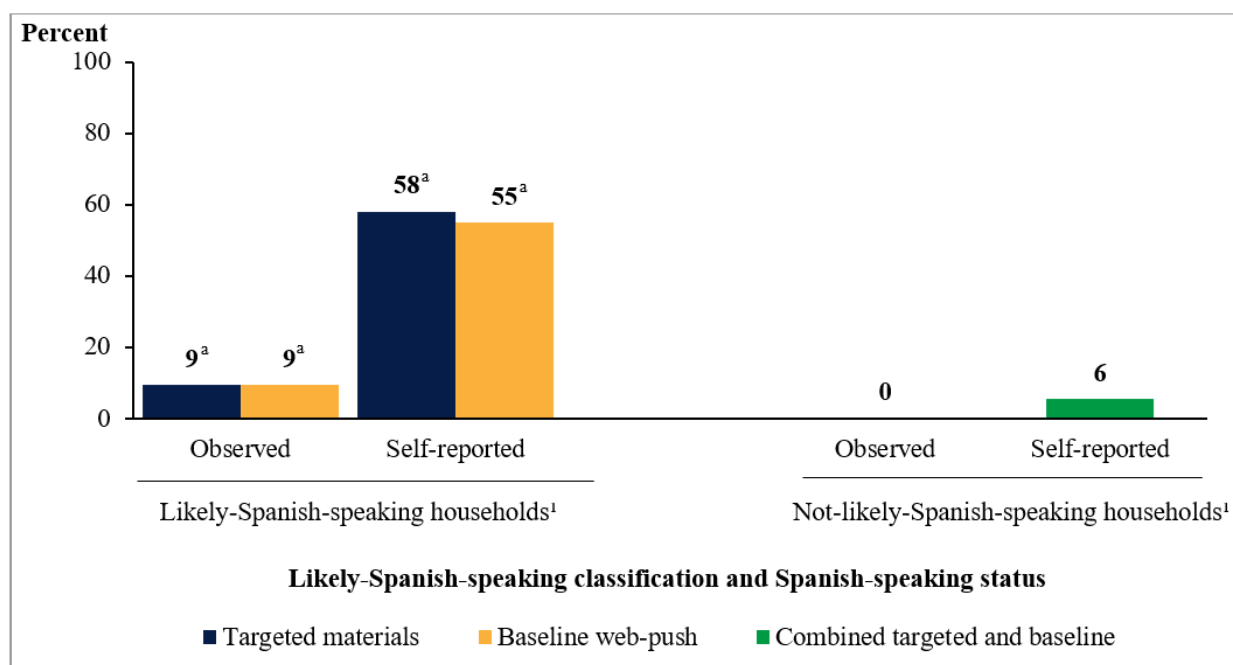
---

<sup>4</sup> Respondents were flagged as self-reported Spanish speakers if any of the following were true: the sampled child was reported as speaking Spanish or English and Spanish equally at home ("What language does this child speak most at home?"); Spanish was reported as being spoken by adults in the household ("What language(s) are spoken at home by the adults in this household?"); parent 1 was reported as speaking Spanish or English and Spanish equally at home ("What language does this parent or guardian speak most at home now?"); or parent 2 (if applicable) was reported as speaking Spanish or English and Spanish equally at home ("What language does this parent or guardian speak most at home now?").

**is no evidence that the tailored materials improved response among actual Spanish-speaking households.**

These results are consistent with both potential explanations for the negative effect of the targeted materials treatment. A substantial proportion of those who received the tailored materials were not, in fact, Spanish-speaking, which could have backfired to the extent that these households perceived the materials as not being relevant to them. However, there is also no evidence that the tailored materials substantially improved response among *actual* Spanish-speaking households.

**Figure 4.2.3. Observed and self-reported Spanish-speaking status, by likely-Spanish-speaking classification and experimental condition: 2019**



<sup>a</sup> Significant difference ( $p < .05$ ) from not-likely-Spanish-speaking households.

<sup>1</sup> Likely-Spanish-speaking households were identified based on information available on the sampling frame: (1) the address is flagged as having a Hispanic surname, (2) the address is in a tract with 40 percent or more Hispanic persons, or (3) the address is in a tract where 10 percent or more of households have Spanish as their primary language and “limited English-speaking” status.

NOTE: Screener respondents were flagged as observed Spanish speakers if they completed the screener in Spanish. Topical respondents were flagged as self-reported Spanish speakers if any of the following were true: the sampled child was reported as speaking Spanish at home; parent 1 was reported as speaking Spanish at home; parent 2 (if applicable) was reported as speaking Spanish at home; or the respondent reported that Spanish was spoken in the home. For observed Spanish speaking status, unweighted sample size was 1,140 for likely-Spanish-speaking cases in the targeted materials condition, 3,710 for likely-Spanish-speaking cases in the baseline web-push condition, and 23,690 for not-likely-Spanish-speaking cases in the targeted materials and baseline web-push conditions. For self-reported Spanish speaking status, unweighted sample size was 300 for likely-Spanish-speaking cases in the targeted materials condition, 980 for likely-Spanish-speaking cases in the baseline web-push condition, and 4,780 for not-likely-Spanish-speaking cases in the targeted materials and baseline web-push conditions. Sample sizes have been rounded to the nearest 10. Detail may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.

## Chapter 5. Summary and Conclusions

This final chapter of the report summarizes the key findings and overarching conclusions for each experiment (see exhibit 5.1 for the screener, topical, and overall response rates for each experiment). It also notes which experimental conditions will be incorporated into the next NHES administration.

**Exhibit 5.1. Screener, topical, and overall response rate, by experimental condition**

Experimental condition	Screener response rate	Topical response rate		Overall response rate	
		ECPP	PFI	ECPP	PFI
Choice-plus experiment					
Choice-plus \$10	64.7	85.5	83.4	55.4	54.0
Choice-plus \$20	66.0	90.1	87.8	59.5	58.0
Baseline web-push	58.9	84.0	81.6	49.5	48.1
Random paper-only	64.3	72.1	73.0	46.4	46.9
Modeled-mode experiment					
Modeled-mode	60.4	83.4	82.1	50.4	49.6
Baseline web-push	58.9	84.0	81.6	49.5	48.1
Random paper-only	64.3	72.1	73.0	46.4	46.9
Advance mailings experiment					
Advance mailing campaign	58.4	87.7	84.5	51.2	49.3
Advance letter-only	58.3	88.8	85.5	51.8	49.9
No-advance mailings	57.2	85.6	85.3	49.0	48.8
Pressure-sealed envelope versus reminder postcard					
Pressure-sealed envelope <sup>1</sup>	37.5	–	–	–	–
Reminder postcard <sup>1</sup>	33.3	–	–	–	–
FedEx timing experiment					
Random FedEx second	58.4	88.9	87.1	51.9	50.8
Random FedEx third	58.9	84.0	81.8	49.5	48.2
Random FedEx fourth	57.4	84.8	83.1	48.7	47.7
Modeled FedEx	58.3	88.6	85.1	51.7	49.7
Opt-out screener materials experiment					
Opt-out screener materials	60.2	89.2	85.6	53.7	51.8
Regular screener materials	58.6	89.8	87.5	52.6	51.3
Targeted Spanish-speaking-household materials experiment					
Targeted materials	58.3	83.7	83.5	48.8	46.7
Baseline web-push	58.9	84.0	81.6	49.5	48.1

–Not applicable. These response rates were not calculated as part of comparing the pressure-sealed envelope and reminder postcard.

<sup>1</sup> This is the response rate after the pressure-sealed envelope or reminder postcard was sent; it is not the final screener response rate.

NOTE: ECPP is Early Childhood Program Participation topical survey. PFI is Parent and Family Involvement topical survey. Response rates were calculated using American Association for Public Opinion Research (AAPOR) Response Rate 1 (RR1). The screener response rate represents the proportion of eligible sampled households (excluding undeliverable and out-of-scope addresses) that were respondents to the screener questionnaire. The topical response rate represents the proportion of eligible households that were respondents to the topical questionnaire. The overall response rate is the product of the screener response rate and the topical response rate.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Household Education Surveys Program (NHES), 2019.



## 5.1 Leveraging Response Mode Assignment to Increase Response Rates

The results of the NHES:2016 web-push experiment suggested that using a web-push design is preferable to paper-only for most sample members. However, it was hypothesized that there may be a cohort of “paper-sensitive” cases that will not respond to the screener unless offered a paper questionnaire from the start of the data collection. As a result, NHES:2019 tested two potential approaches to reducing the impact of paper sensitivity on response rates while still obtaining web responses from as many cases as possible.

### 5.1.1 Choice-Plus Experiment

In the choice-plus condition, sample members were offered web, paper, and inbound telephone response options in all screener packages. They also were offered a promised incentive for responding by web or by inbound telephone; most were offered \$10, but some were randomly assigned to be offered \$20.

**Effect of the choice-plus protocol:** Among all the experimental conditions that were tested in NHES:2019, the choice-plus conditions were the most successful at maximizing both the screener and overall response rates.

- Both the \$10 and \$20 choice-plus conditions resulted in screener response rates that were higher than in the baseline web-push condition (by 6 and 7 percentage points, respectively). They led to particularly large increases in the response to early screener mailings compared to the baseline web-push condition (by 15 and 17 percentage points, respectively). There was some evidence that the choice-plus conditions were particularly effective at increasing screener response among households with lower-than-average screener response rates.
- The overall response rates for both the ECPP and PFI topical surveys also were higher in the choice-plus conditions than in the baseline web-push or random paper-only conditions (by at least 6 percentage points).

**Effect of the larger incentive value:** The \$20 choice-plus condition was somewhat more successful than the \$10 choice-plus condition at maximizing overall response rates.

- Though the screener response rate was not significantly different in the two conditions, the \$20 incentive led to a higher screener response rate *by web* (by 4 percentage points) as compared to the \$10 incentive. Because web screener respondents tend to have higher topical response rates than paper screener respondents, this led to the \$20 condition having higher topical response rates and overall response rates than the \$10 condition for both topical surveys (by 4 to 5 percentage points).

**Use of choice-plus in the next NHES administration:** The response rate findings make choice-plus an attractive option for use in future NHES administrations. However, cost analyses conducted by the Census Bureau (the NHES data collector) suggest that—despite the higher choice-plus response rates—the use of choice-plus for the entire NHES sample would be cost prohibitive (Varela and Zotti 2020). For example, the cost per screener response in the \$20 choice-plus condition was 1.35 times the cost in the baseline web-push condition, and the cost per topical response was 1.17 times the cost in the baseline

web-push condition. This is likely due to the addition of the promised incentive and the increased use of paper questionnaires in the choice-plus conditions. As a result, the next NHES administration will use a \$20 choice-plus design for lower propensity cases to maximize response rates among these cases while also constraining costs.

### 5.1.2 Modeled-Mode Experiment

In the modeled-mode condition, sample members who were expected to be paper-sensitive (that is, based on a predictive model, were expected to be much more likely to respond under a paper-only design than a web-push design) were offered a paper questionnaire in all mailings and were not offered the option to respond by web. All other sample members in this condition received the baseline web-push protocol.

**Effect of the modeled-mode approach over the conditions as a whole:** The screener and overall response rates in the modeled-mode condition were relatively similar to those achieved in the baseline web-push and random paper-only conditions.

- The modeled-mode condition resulted in a screener response rate that was slightly higher than in the baseline web-push condition (by 2 percentage points) but lower than in the random paper-only condition (by 4 percentage points).
- The PFI overall response rate was higher in the modeled-mode condition than in the baseline web-push condition (by 2 percentage points). However, there were no other statistically significant differences between the overall response rates for the modeled-mode condition and the other two conditions.

**Ability to identify paper-sensitive cases:** The predictive model was successful at identifying cases whose response status after the first two screener packages was more sensitive than average to the response mode offered in those packages. But it was not successful at identifying a group of cases whose final screener response status was more sensitive than average to the sequence of modes offered over all mailings.<sup>1</sup>

**Use of a modeled-mode approach in the next NHES administration:** As a result of the response rate outcomes discussed above and the increased operational complexity of implementing a targeted design, the next NHES administration will not include a modeled-mode approach.

## 5.2 Refining Web-Push Contact Procedures

In an effort to improve upon the success of the NHES:2016 web-push experiment, a subset of the NHES:2019 sample was randomly assigned to an updated web-push mailing protocol whose goal was to

---

<sup>1</sup> The goal of this experiment was to identify a group for which the screener response rate increase with a paper-only protocol (versus a web-push one) was larger than the increase for non-paper-sensitive cases. As shown in table D.1 in appendix D, the final screener response rate was higher with a paper-only protocol than with a web-push protocol for **both** paper-sensitive cases and non-paper-sensitive cases. However, the increase among paper-sensitive cases was **not** significantly greater than among non-paper-sensitive cases.

further refine the NHES web-push methodology. There were several experiments conducted as part of this protocol, which are described in greater detail in the remainder of this section.

### 5.2.1 Advance Mailings Experiment

Sample members were assigned to one of three advance mailing conditions: advance letter only, advance mailing campaign, or no advance mailings. In the advance mailing campaign condition, sample members were sent two oversized postcards prior to the advance letter that displayed NHES:2016 findings but did not mention that the address had been sampled for NHES:2019; the purpose of the postcards was to build familiarity and engagement with the NHES.

**Effectiveness of the advance letter:** While sending an advance letter did not increase the final screener response rate, it did increase early screener response (by 4 percentage points) and screener response by web (by 3 percentage points). This suggests that the advance letter helped reduce screener-phase costs by decreasing the number of follow-up mailings that needed to be sent and the number of paper questionnaire responses that needed to be processed. Sending an advance letter also increased the ECPP overall response rate (by 3 percentage points) but did not have a significant impact on the PFI overall response rate.

**Effectiveness of the advance mailing campaign:** The advance mailing campaign had no effect on the screener or overall response rates beyond sending an advance letter only.

**Use of advance mailings in the next NHES administration:** The next NHES administration will include an advance letter but will not incorporate the advance mailing campaign postcards.

### 5.2.2 Pressure-Sealed Envelope Versus Reminder Postcard

Sample members who received the updated web-push mailing protocol were sent a pressure-sealed envelope after the initial screener package, while sample members in the baseline web-push condition were sent a reminder postcard. The purpose of both materials was to remind sample members to respond who had not yet done so and to thank those who had already responded. The pressure-sealed envelope included the web survey URL and the household's unique web login credentials, but the reminder postcard did not (because the postcard format did not allow for sufficient protection of the household's web login credentials).

**Effectiveness of the pressure-sealed envelope:** Sending a pressure-sealed envelope increased response to that reminder mailing by 4 percentage points as compared to sending a reminder postcard.<sup>2</sup>

**Use of pressure-sealed envelopes/reminder postcards in the next NHES administration:** The next NHES administration will use a pressure-sealed envelope as a reminder after the initial screener mailing, rather than a reminder postcard.

---

<sup>2</sup> As a result of differences in the subsequent mailings sent in these conditions, it was not possible to isolate the effect of the pressure-sealed envelope on the final screener response rate.

### 5.2.3 FedEx Timing Experiment

The goal of the FedEx timing experiment was to determine the ideal timing for sending FedEx reminders in a web-push design, balancing the potential response gains of sending FedEx mailings earlier with the additional shipping costs associated with doing so. Sample members were assigned to one of three conditions: random FedEx second (in which all sample members were sent the second screener package via FedEx), random FedEx fourth (in which all sample members were sent the fourth screener package via FedEx), and modeled FedEx (in which FedEx-high-priority cases were sent the second screener package via FedEx and FedEx-low-priority cases were sent the fourth screener package via FedEx). These conditions also were compared to the baseline web-push condition, in which sample members were sent the third screener package via FedEx (in the context of this experiment, referred to as the “random FedEx third” condition).

**Ability to identify FedEx-low-priority cases:** The model was successful at identifying cases that were FedEx-low-priority. Delaying the FedEx mailing for those cases did not reduce their final screener response rate. But sending the FedEx mailing earlier for FedEx-high-priority cases did increase their final screener response rate (by 2 percentage points).

**Effect of FedEx over the conditions as a whole:** Although sending the FedEx mailing earlier in the data collection did not have a large effect on the final screener response rate, it increased response to earlier mailings and increased the screener response rate by web—both of which had a positive impact on the overall response rates.

- The final screener response rate in the random FedEx fourth condition was lower than in the random FedEx second, random FedEx third, and modeled FedEx conditions (by 1 to 2 percentage points). But there were no statistically significant differences between the final screener response rates for the other three conditions.
- However, in the random FedEx second condition, the screener response rate after the early mailings was higher than in the random FedEx third and FedEx fourth conditions (by 7 percentage points), suggesting that sending the FedEx mailing earlier helped reduce the number of follow-up mailings that needed to be sent and the number of paper questionnaire responses that needed to be processed.
- The screener response rate by web also was higher in the random FedEx second condition than in the random FedEx third or fourth conditions (by 5 to 6 percentage points). Because web screener respondents tend to have higher topical response rates than paper screener respondents, this led to the overall response rates for both topical surveys being higher in the random FedEx second condition than in the random FedEx third or fourth conditions (by 2 to 3 percentage points).
- The modeled FedEx results tended to fall between those of the random FedEx second and random FedEx fourth conditions. Overall, they were most similar to the random FedEx second condition (likely because most of the modeled FedEx condition cases were sent the second screener package via FedEx).

- The cost per response in the modeled FedEx condition was slightly lower than in the random FedEx second condition. However, there do not appear to be sufficient cost savings associated with modeling FedEx timing to make it worth the increased operational complexity.

**FedEx timing for the next NHES administration:** The next NHES administration will send the second screener package via FedEx for all cases. Using FedEx at the second screener package helps to increase screener response by web, which in turn helps to increase the overall response rates. Although the response rate results were relatively similar for the random FedEx second and modeled FedEx conditions, delaying the FedEx mailing for some cases in the modeled condition did not result in much of a cost savings and would increase the operational complexity of the data collection.

## 5.3 Attracting Response From Specific Demographic Subgroups

The final two experiments tested the use of mailings that were specially designed to improve the screener response rates among two key subgroups within the NHES:2019 sample. The goal of the opt-out screener materials experiment was to test survey materials that would encourage households without NHES-eligible children to respond to the screener. The targeted Spanish-speaking-household materials experiment sought to determine whether tailoring the wording and design of bilingual materials could improve the response rates of this critical subgroup. However, neither of these treatments increased the screener or overall response rates among the targeted subgroups. These results speak to the potential challenges of using targeted approaches in address-based samples, where the information available about sampled households prior to data collection may be limited or inaccurate.

### 5.3.1 Opt-Out Screener Materials Experiment

In the opt-out screener materials condition, the screener contact materials were adapted to help convey to households without children that this is a low-burden request, and the paper screener questionnaire was modified so that the first item asking whether the household had any children was on the front cover of the screener booklet (allowing households without children to immediately see that they only needed to respond to that item).

**Effect of opt-out screener materials on households without children:** The results of this experiment were the opposite of expectations.

- Among households not flagged on the NHES sampling frame as having children, the opt-out materials had no impact on the screener response rate. However, among households that were flagged as having children, the opt-out materials increased the screener response rate (by 5 percentage points).
- A possible explanation for this unexpected pattern is that the opt-out wording may have primed *all* households, regardless of whether they have children, to expect the survey to be low burden. In this way, the opt-out wording could have encouraged greater response from the group most likely to find the survey salient (households with children). An alternate explanation may be that this opt-out wording made it clearer that the survey primarily focused on households with children, again making the screener more salient for households with children and less salient for households without children.

- Prior to data collection, the “presence of children” flag available on the sampling frame is the best available indicator of whether there are NHES-eligible children living at the sampled addresses. However, in NHES:2019, it both captured a significant number of addresses that did not have NHES-eligible children and missed a considerable number of addresses that did have them. Consequently, if this flag were used to target opt-out materials in future administrations, there would be substantial inaccuracy in the targeting.

**Effect of opt-out screener materials over the conditions as a whole:** The opt-out materials did not have a statistically significant or practically meaningful impact on the screener or overall response rates.

**Use of opt-out screener materials in the next NHES administration:** Due to the lack of an effect on response among households without children or on response overall, the next NHES administration will not use the opt-out screener materials. Other strategies for communicating the purpose of the screener to households with children and to households without children are being considered.

### 5.3.2 Targeted Spanish-Speaking-Household Materials Experiment

In this condition, likely-Spanish-speaking households (identified based on auxiliary data available on the sampling frame or appended from publicly available sources) were sent tailored materials that were expected to improve response among Spanish-speaking households. All other households in this condition were sent the standard NHES web-push materials.

**Effect of tailored materials on likely-Spanish-speaking households:** The tailored materials had a negative effect on the screener and overall response rates among likely-Spanish-speaking households.

- As seen in prior administrations, when using the standard NHES materials, the screener response rate was lower among likely-Spanish-speaking households than among not-likely-Spanish-speaking households (by 15 percentage points). The tailored materials further decreased the screener response rate among likely-Spanish-speaking households compared to the baseline mailings (by 6 percentage points). They also decreased the ECPP and PFI overall response rates among likely-Spanish-speaking households (by 6 percentage points).
- A potential explanation for this negative finding is that the tailored materials backfired among households that were flagged as likely-Spanish-speaking but were not, in fact, Hispanic or Spanish-speaking. The likely-Spanish-speaking indicator used in this experiment also appeared to be relatively imprecise; among likely-Spanish-speaking cases in the targeted condition, only 9 percent of respondents completed the screener in Spanish and only 58 percent of topical respondents reported speaking some Spanish at home. An alternate explanation is that the materials themselves inadvertently discouraged response by suggesting that the government was aware of individual sample members’ ethnic heritage or language status, which may have then made Hispanic or Spanish-speaking sample members more reluctant to participate in the survey.

**Effect of tailored materials over the conditions as a whole:** Because likely-Spanish-speaking households made up a relatively small share of the sample, the impact of the tailored materials on the screener and overall response rates over the full sample was not statistically significant.

**Use of tailored Spanish-speaking-household materials in the next NHES administration:** Due to their negative effect on the screener and overall response rates among likely-Spanish-speaking households, the next NHES administration will not use the tailored materials.

## References

- American Association for Public Opinion Research. (2016). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th ed.). AAPOR.
- Biemer, P.P., Murphy, J., Zimmer, S., Berry, C., Deng, G., and Lewis, K. (2018). Using Bonus Monetary Incentives to Encourage Web Response in Mixed-Mode Household Surveys. *Journal of Survey Statistics and Methodology*, 6(2): 240-61.
- Dillman, D.A., Smyth, J., and Christian, L.M. (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (4th ed.). Wiley.
- Evans, S., Levy, J., Miller-Gonzalez, J., Vines, M., Girón, A.S., Walejko, G., Bates, N., and Trejo, Y.G. (2019). *2020 Census Barriers, Attitudes, and Motivators Study (CBAMS) Focus Group Final Report: A New Design for the 21st Century*. U.S. Department of Commerce. Washington, DC: U.S. Census Bureau. Retrieved January 19, 2021, from <https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2020-report-cbams-focus-group.pdf>.
- Groves, R.M., Presser, S., and Dipko, S. (2004). The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly*, 68(1): 2-31.
- Jackson, M., Kaiser, A., Battle, D., Wan, C., Quenneville, G., Kincel, B., and Cox, C. (2021). *National Household Education Surveys Program of 2019: Data File User's Manual* (NCES 2021-030). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved February 1, 2021, from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2021030>.
- Jackson, M.T., Medway, R.L., and Megra, M.W. (2021). Can Appended Auxiliary Data Be Used to Tailor the Offered Response Mode in Cross-Sectional Studies? Evidence From an Address-Based Sample. *Journal of Survey Statistics and Methodology* (Advance online publication). <https://doi.org/10.1093/jssam/smab023>
- McPhee, C., Jackson, M., Bielick, S., Masterton, M., Battle, D., McQuiggan, M., Payri, M., Cox, C., and Medway, R. (2018). *National Household Education Surveys Program of 2016: Data File User's Manual* (NCES 2018-100). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Medway, R., Noel, H., and Guarino, N. (2018). *NHES:2017 Web Test Report* (Unpublished report). U.S. Department of Education, National Center for Education Statistics.
- Medway, R., Scardaville, M., Paek, C., Dias, P., Kaiser, A., Megra, M., and Pulizzi, S. (2022). *National Household Education Surveys Program of 2019: Study of Nonresponding Households* (NCES 2022-043). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved February 1, 2022, from <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2022043>.



- National Institute of Statistical Sciences (NISS). (2018). *NCES Study and Survey Recruitment Planning and Materials*. Retrieved January 19, 2021, from [https://www.niss.org/sites/default/files/research\\_attachments/Study%20%26%20Survey%20Recruitment%20Planning%20%26%20Materials-FT.pdf](https://www.niss.org/sites/default/files/research_attachments/Study%20%26%20Survey%20Recruitment%20Planning%20%26%20Materials-FT.pdf).
- Office of Management and Budget. (2018). *National Household Education Survey 2019 (NHES:2019) Information Collection Request: Appendix 5 - NHES 2019 Developmental Studies Results*. Retrieved February 25, 2021, from <https://www.reginfo.gov/public/do/DownloadDocument?objectID=89547801>.
- Roth, S.B., Han, D., and Montaquila, J. (2013). The ABS Frame: Quality and Considerations. *Survey Practice*, 6(4). <https://doi.org/10.29115/SP-2013-0021>
- Varela, K., and Zotti, A. (2020). *NHES Cost Analysis for the Experimental Treatment Groups in the 2016 and 2019 Data Collection Cycles* (Unpublished memo). Prepared by the Census Bureau for the U.S. Department of Education, National Center for Education Statistics.
- West, B.T., Wagner, J., Hubbard, F. and Gu, H. (2015). The Utility of Alternative Commercial Data Sources for Survey Operations and Estimation: Evidence From the National Survey of Family Growth. *Journal of Survey Statistics and Methodology*, 3(2): 240-64.