*Journal of Survey Statistics and Methodology (2022) **00**, 1–17*

AN EXPERIMENTAL EVALUATION OF TWO APPROACHES FOR IMPROVING RESPONSE TO HOUSEHOLD SCREENING EFFORTS IN NATIONAL MAIL/WEB SURVEYS

JAMES WAGNER*
BRADY T. WEST
MICK P. COUPER
SHIYU ZHANG
REBECCA GATWARD
RAPHAEL NISHIMURA
HTAY-WAH SAW

Survey researchers have carefully modified their data collection operations for various reasons, including the rising costs of data collection and the ongoing Coronavirus disease (COVID-19) pandemic, both of which have made in-person interviewing difficult. For large national surveys that require household (HH) screening to determine survey eligibility, cost-

JAMES WAGNER is a Research Professor, BRADY T. WEST is a Research Associate Professor, MICK P. COUPER is a Research Professor, SHIYU ZHANG is a PhD Student in the Michigan Program in Survey and Data Science, REBECCA GATWARD is a Survey Director, RAPHAEL NISHIMURA is a Sampling Operations Director, and HTAY-WAH SAW is a PhD Student in the Michigan Program in Survey and Data Science with the Survey Research Center, Institute for Social Research, University of Michigan, USA.

This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health (grant number R01HD095920). We also thank Colette Keyser, Andrew Hupp, and the rest of the AFHS team. This study's design and analysis were not preregistered.

The National Survey of Family Growth (NSFG) is conducted by the Centers for Disease Control and Prevention's (CDC's) National Center for Health Statistics (NCHS). The data collection included in this article was performed under contract # 200-2010-33976 with the University of Michigan's Institute for Social Research with funding from several agencies of the U.S. Department of Health and Human Services, including CDC/NCHS, the National Institute of Child Health and Human Development (NICHD), the Office of Population Affairs (OPA), and others listed on the NSFG webpage (see http://www.cdc.gov/nchs/nsfg/). The views expressed here do not represent those of NCHS or the other funding agencies.

*Address correspondence to James Wagner, University of Michigan, 4053 ISR, 426 Thompson St., Ann Arbor, MI 48104, USA; E-mail: jameswag@umich.edu.

https://doi.org/10.1093/jssam/smac024
© The Author(s) 2022. Published by Oxford University Press on behalf of the American Association for Public Opinion Research. All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

efficient screening methods that do not include in-person visits need additional evaluation and testing. A new study, known as the American Family Health Study (AFHS), recently initiated data collection with a national probability sample, using a sequential mixed-mode mail/web protocol for push-to-web US HH screening (targeting persons aged 18–49 years). To better understand optimal approaches for this type of national screening effort, we embedded two randomized experiments in the AFHS data collection. The first tested the use of bilingual respondent materials where mailed invitations to the screener were sent in both English and Spanish to 50 percent of addresses with a high predicted likelihood of having a Spanish speaker and 10 percent of all other addresses. We found that the bilingual approach did not increase the response rate of high-likelihood Spanish-speaking addresses, but consistent with prior work, it increased the proportion of eligible Hispanic respondents identified among completed screeners, especially among addresses predicted to have a high likelihood of having Spanish speakers. The second tested a form of nonresponse follow-up, where a subsample of active sampled HHs that had not yet responded to the screening invitations was sent a priority mailing with a $5 incentive, adding to the $2 incentive provided for all sampled HHs in the initial screening invitation. We found this approach to be quite valuable for increasing the screening survey response rate.

KEYWORDS: ABS sample; Incentives; Nonresponse.

Statement of Significance

Many surveys are moving from interviewer- to self-administered modes. The relatively high cost of conducting interviewer-administered modes is one important reason for this move. As surveys make this transition, they will need to identify effective protocols that are aimed at recruiting respondents from many different backgrounds. The research in this article identifies methods that improve both response rates and data quality. The latter is accomplished by identifying methods that improve response among groups that are otherwise low-responding. This research identifies tools that researchers conducting relatively low-cost surveys will need to achieve the highest quality estimates possible.

1. INTRODUCTION

Survey researchers have carefully modified their data collection operations in response to the ongoing COVID-19 pandemic, which has made in-person interviewing difficult (if not impossible) in most countries. Even before the pandemic, the costs of in-person interviewing were rising while response rates

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

were declining (Williams and Brick 2018). For large national surveys that require some form of household (HH) screening to determine survey eligibility, the problem is particularly complex. Face-to-face surveys with low eligibility rates may require visits to many HHs that will turn out to be ineligible. Alternative screening methods that do not require in-person visits and can improve cost efficiency are sorely needed, but such new approaches require an additional and rigorous evaluation and testing (Wagner, Arrieta, Guyer, and Ofstedal 2014; Biemer, Murphy, Zimmer, Berry, Deng, et al. 2018; DeBell, Amsbary, Meldener, Brock, and Maisel 2018a; DeBell, Jackman, Maisel, Amsbary, Meldener, et al. 2018b).

As part of a new study known as the American Family Health Study (AFHS; see afhs.isr.umich.edu), we recently initiated data collection with a national address-based probability sample of more than 19,000 US addresses. The AFHS used a sequential mixed-mode mail/web protocol for push-to-web HH screening to identify eligible persons aged 18–49. To better understand optimal approaches for this type of national mail/web screening effort, we embedded two randomized experiments in the AFHS data collection:

(1) Mailed screener invitations (and paper screener questionnaires in subsequent mailings to non-responders) were sent in English *and* Spanish to a random 50 percent of addresses with a high predicted likelihood of having a Spanish speaker, and a random 10 percent of all other addresses. The balance of the sample was sent materials in English only.
(2) A random subsample of addresses that had received (1) an initial mailed invitation to complete the screening survey, (2) a postcard with a reminder to complete the screening survey, and (3) a mailed paper screener but had not yet responded, was sent a priority mailing with a $5 incentive in the last phase, adding to the $2 incentive provided for all sampled HHs. No additional effort was applied to the other nonrespondents that were not part of this random subsample.

We present the results of these two experiments and discuss useful future directions for improving the cost efficiency of screening operations in large-scale web/mail data collections that do not involve in-person interviews.

2. BACKGROUND

2.1 Recruitment of Spanish-Speaking Households

The effect of bilingual (Spanish and English) materials on response has been investigated experimentally, including in mailed screening surveys (and, in some cases, mailed bilingual screening instruments). In one of the earliest known studies, Bouffard and Tancreto (2006) report on a test done in 2005 to improve response rates for the 2010 Census using a mail instrument. A random

sample of addresses was mailed a bilingual form, with the rest being sent an English-only form. Mail response rates were 2.2 percentage points higher for the bilingual form, a marginally significant difference ($p < .10$). They found a significant increase in response for the bilingual form in the high non-White or Hispanic concentration stratum (3.2 percentage point increase; S.E. = 0.81), as well as a significant increase (1.7 percentage point increase; S.E. = 0.79) in areas with a low concentration of the non-White and Hispanic populations, suggesting that concerns about a "backfire effect"—lower response rates among English speakers because of the inclusion of Spanish materials (see Brick, Caporaso, Williams, and Cantor 2019)—were not supported by available data.

Brick, Montaquila, Han, and Williams (2012) report on a series of experiments designed to improve response rates from Spanish speakers in the 2011 National Household Education Surveys (NHES) Program. They used two methods to identify likely Spanish-speaking HHs: 1) Census tracts with at least 13 percent of HHs classified as linguistically isolated Spanish speakers and 2) addresses with a surname that was likely to be of Hispanic origin. They tested bilingual and English-only materials in a variety of different combinations. They found that sending survey materials to all sampled HHs in both languages substantially increased participation from Spanish-speaking HHs and did not jeopardize responses from monolingual English HHs. They explain their findings in terms of accommodation theory (Koslow, Shamdasani, and Touchstone 1994), social exchange theory (Dillman, Smyth, and Christian 2014) or the norm of reciprocity (Groves, Cialdini, and Couper 1992), all of which suggest that supporting group values in the request increases survey response.

In a more recent study, Elliott, Klein, Kallaur, Brown, Hays, et al. (2019) used surname and address to predict Spanish-language preference for a large national sample of Medicare beneficiaries. They randomized half of the 10,000 non-Puerto Rico beneficiaries with the highest predicted probabilities of Spanish preference (>10 percent probability) to bilingual mailings (intervention) and a half to standard English-only mailings (control). They found significantly higher ($p < .0001$) mail response rates for the intervention (28.7 percent percent) than the control (23.9 percent). In comparing the experimental treatments, they estimated that targeted bilingual mailings induced 6.5 percent of those who would not have responded to respond by mail and 54.0 percent of those who would have responded in English to respond in Spanish. Beneficiaries more likely to be Spanish showed greater increases in response rates, a higher proportion of responses in Spanish, and lower response rates in the control condition. They concluded that targeted bilingual mailings using commonly-available surname and address information can efficiently increase the participation of Hispanics in surveys.

Collectively, these findings suggest that bilingual mailings have a beneficial effect on mail response rates in areas with high concentrations of Hispanics while having little or no deleterious effects in predominantly English-speaking

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

areas. However, given that an early concern raised about this approach is the potential for a "backfire effect," we used a disproportionate allocation of the bilingual materials in the present study, sending them to 50 percent of addresses predicted to have a high likelihood of speaking Spanish and a random 10 percent of all other addresses to assess this claim.

2.2 The Use of Incentives in Mixed-Mode Surveys

The use of small unconditional incentives in mail or mixed-mode survey invitations is well-established. Social exchange theory (Dillman 1978) or leverage-saliency theory (Groves, Singer, and Corning 2000) are often used to justify such incentives. Evidence for the effectiveness of such incentives is documented in meta-analyses (e.g., Church 1993; Edwards, Roberts, Clarke, DiGuiseppi, Pratap, et al. 2002; Mercer, Caporaso, Cantor, and Townsend 2015). In general, token cash incentives are typically used only once in mail surveys in the initial invitation (Messer and Dillman 2011). Given this, the literature on using additional incentives in follow-up mailings to nonrespondents in mail surveys is less developed (see Dykema, Jaques, Cyffka, Assad, Hammers, et al. 2015; Dirksz, Lavrakas, Lusskin, Ponce, and Felstead 2017), as is the theoretical motivation for doing so.

Some surveys have experimented with the use of an incentive offer as part of a nonresponse follow-up strategy. Kropf and Blair (2005) conducted a small experiment among sample persons (members of health maintenance organizations in Maryland) who did not respond to an initial invitation (with no incentive) to complete a mail survey. They varied both the content of the cover letter and the incentive ($5 versus none). They found that adding the $5 incentive increased response by more than 20 percentage points. Messer and Dillman (2011) experimented with methods to push sample members to complete surveys online using postal mailings, including a set of experiments in which nonresponders were sent an additional incentive in a priority mailing. They found that sending a second $5 incentive in a priority mailing significantly increased response rates by 9 percentage points, in contrast to sending either the initial $5 incentive only, which all sample groups were sent, or the follow-up priority mailing only. However, they did not include a condition with an incentive but without priority mailing, so it is not possible to disentangle the effects of the second incentive from the format of the mailing.

Dykema et al. (2015) report on studies to examine the effects of sequential prepaid incentives and envelope messaging in two mail surveys using address-based samples. Sample members in the first study were randomly assigned to groups that received an initial incentive of $2 versus $5, and the second incentive of $0 versus $2 (sent only to nonresponders). Results indicated that the $5 initial incentive significantly increased response rates, but the second incentive had no additional effect. In the second study, they delivered a $2 initial

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

incentive to all sample members but increased the amount of the second incentive to $5. They found that the second incentive, valued at a larger amount than the initial incentive, significantly increased response rates by 6 percentage points.

Dirksz et al. (2017) report on an experiment of an additional $1 incentive (versus $0) in a follow-up mailing to nonrespondents in a national mail survey. They tested the additional incentive in two separate strata, a non-Hispanic sample and a Hispanic sample. The incentive significantly raised the response rate among nonrespondents in the non-Hispanic sample (from 2.4 to 4.1 percent; $p < .02$) but not in the Hispanic sample (from 2.1 to 4.2 percent; $p > .10$).

Bucks, Couper, and Fulford (2020) report on an experiment in mailed recruitment to a mixed-mode (web or mail) survey in a national sample drawn from credit bureau data. They experimentally varied two factors: (1) using a concurrent web/mail or a sequential web and then mail design, and (2) different incentive amounts ($5 versus $10) given to nonrespondents to the initial survey (which included no incentive). They found no significant differences when comparing response rates across the two incentive amounts.

In terms of costs, Dykema et al. (2015) found that the second $5 incentive did increase costs per interview, but only by an additional $2.59. Messer and Dillman (2011) found that costs per interview did increase (from $37.82 to $47.42, an increase of $9.60) with the additional $5 incentive and priority mailing. In sum, these results are mixed, but generally suggest that a second incentive offered to nonrespondents to initial recruitment efforts can be effective in increasing response rates. These effects may depend upon the incentive amount, the format of the mailing (priority mailing versus standard mailing), and other factors such as the topic of the survey, the length of the survey, and the target population.

3. METHODS

3.1 Survey Design

The AFHS sample studied here is a national address-based probability sample of 19,381 US addresses. A sequential mixed-mode protocol including web and mail screening surveys was implemented to identify HHs eligible for a "main" survey focused on family formation, reproduction and other demographic topics. The screening survey, completed by a HH informant, collected a list of persons in the HH along with their ages to determine eligibility (18–49 years of age).

Figure 1 shows the sequence of contact attempts involved in the recruitment protocol for the screening survey. In the first phase of this protocol, a stratified probability sample of addresses (with oversampling of addresses predicted to be likely eligible and high-density minority areas) received a mailed invitation

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

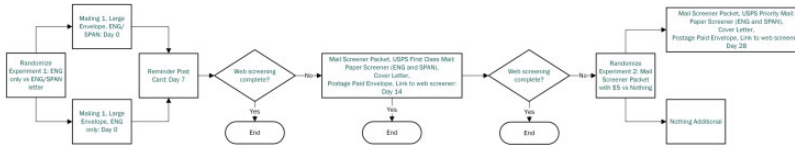

Figure 1. AFHS Screening Survey Protocol, including Experiments 1 and 2.

(including a $2 incentive) to complete a screening questionnaire online (see supplemental materials). In the second phase of this protocol, a follow-up reminder was sent one week after the mailed invitation in the form of a postcard to those who had not responded to the initial mailed invitation (see supplemental materials). In the third phase, a follow-up mailing with a paper version of the screening questionnaire was sent one week after the postcard. Information obtained from completed screening questionnaires was used to identify eligible persons within the sampled HHs. If there was only one eligible person in the HH, then that person was invited to complete the main AFHS survey online. If there was more than one eligible person, one person was randomly selected to complete the main survey. Figure 1 provides an overview of the screening data collection protocol, including the two experiments.

Once an eligible respondent was randomly selected from a sampled HH completing the screening questionnaire, an initial invitation to complete the main survey online was sent by mail to the selected respondent, and the letter promised a $70 token of appreciation once the completed survey was received. This initial invitation was followed two weeks later by either a postcard or email reminder (if the selected respondent provided an email address). After three and five weeks a further reminder was sent by email or text message (when these contact details were provided). For eligible nonrespondents for whom we did not have an email or text-enabled phone number, we mailed a follow-up at four and six weeks that included a substantially shortened paper version of the questionnaire but still encouraged the respondent to complete the survey online. The six-week reminder was mailed in a USPS priority mailer.

After six weeks, our calling center staff made reminder telephone calls to nonrespondents with telephone numbers available from either commercial data sources linked to our sampling frame or the initial screening questionnaire [83 percent of these nonrespondents had telephone numbers available—although some (12 percent) of these were found to be invalid during the reminder calls]. These staff did not administer the survey, but rather encouraged the nonrespondents to self-administer the survey and provided any information that would assist them in doing so. Additional details regarding the AFHS sample design and data collection methodology can be found elsewhere (https://afhs.isr.umich.edu/about-the-study/afhs-methodology/).

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

3.2 Experiment 1

For the first experiment, we used predicted probabilities of a "Spanish Speaker in Household" (defined below) to create two strata—predicted "High" and "Low" likelihood of someone within the HH-speaking Spanish. Then we randomized cases within each stratum to receive either bilingual (Spanish and English) or monolingual (English) materials. Our approach is similar to that of Elliott et al. (2019). The predicted probabilities were obtained by modeling the probability of a "Spanish Speaker in Household" using data from the National Survey of Family Growth (see https://www.cdc.gov/nchs/nsfg/index.htm for additional details on the NSFG), a large national survey that screens for persons of ages 15–49 and conducts interviews in English and Spanish. The model used auxiliary variables available on the sampling frame and from the commercial data. The binary outcome—"Spanish speaker"—was coded from whether either a screening interview or a follow-up interview was conducted in Spanish. Predictors in these models included Census division, sociodemographic domain of the geographic area (based on percent Hispanic and percent African-American), an indicator of someone in the HH having a Hispanic surname from commercial data, and American Community Survey (ACS) variables measured at the Census block group level, including the following:

- The percentage of all ACS occupied housing units where a Spanish or Spanish Creole language was assigned as the HH language, and no one aged 14 years and over speaks English only or speaks English "very well."
- The number of people who identify as "Mexican," "Puerto Rican," "Cuban," or "another Hispanic, Latino, or Spanish origin."
- The number of people who moved from another residence in the United States or Puerto Rico within the last year.

The final estimated coefficients in the fitted logit model are presented in table A.1. Based on an ROC curve analysis of the fitted model (AUC = 0.87), we selected 0.25 as a threshold for the predicted probability p, above which we classified a HH as likely to speak Spanish. Benchmarked against the NSFG survey data, this threshold resulted in an 85.9 percent accuracy rate, a 10.0 percent false-positive rate, and a 32.4 percent false-negative rate.

In the "High Likelihood of a Spanish Speaker" stratum ($n = 2{,}914$), a random 50 percent of cases received bilingual (Spanish and English) materials at each of the three mailings, including the paper screener that was part of the third mailing. The other 50 percent of cases received monolingual (English-only) materials. In the "Low Likelihood of a Spanish Speaker" stratum ($n = 16{,}467$), a random 10 percent of cases received the bilingual materials. The other 90 percent of cases received monolingual materials.

For experiment 1, we compare AAPOR RR1 (AAPOR 2016) response rates for the screening survey between the two treatments (monolingual versus

bilingual materials). We also examine these response rates for sampling domains and strata defined by the likelihood of Spanish being spoken in the HH.

3.3 Experiment 2

In the second experiment, we sent an additional $5 incentive (on top of the $2 incentive that all cases received in the initial mailing) to a random subsample of active cases ($n = 4{,}830$) that had not yet responded after the full screening protocol. The experimental treatment assignment for the second experiment was stratified by the treatment assignment for the first experiment in a full factorial design. This sample size was chosen to be large enough to detect expected effect sizes. Given the cost of the intervention, we wanted to be sure it would be useful before implementing it on a larger group. At the point that this intervention was implemented, 1,706 screener lines had been resolved and the screening response rate was about 7 percent. We sent the incentive in a priority mailer.

The other remaining active cases ($n = 12{,}649$) did not receive additional mailings or incentives. The control condition for this experiment is the group of sampled cases that was still active at the time of the intervention but was not randomized to receive the intervention at the time that the intervention was implemented (i.e., the 12,649) cases. Therefore, the experimental comparison is a priority mailer with an incentive versus no mailing at all.

For the second experiment, we begin with an initial comparison of AAPOR RR1 (AAPOR 2016) response rates to the screening survey between (1) the subsample given the priority mailings with $5 and (2) the remaining active cases. We compare conditional screening response rates between these two groups. We also compare the characteristics of persons responding to the two treatments. We then proceed to estimate a large multivariable logit model, looking at the ability of this "final mailing" intervention to increase screener response propensity when adjusting for a wide variety of auxiliary information available on the sampling frame, including commercial data and US Census Planning Database (PDB) data. Model estimates are provided in table A.2. Finally, we test interactions between receiving the "final mailing" intervention and all of the auxiliary variables included in the model.

Since this is a screening survey, no partial completes were accepted. Undeliverable returned mail was treated as non-sample. We also examine the characteristics of respondents in each of the treatments.

4. RESULTS

4.1 Experiment 1

Table 1 presents AAPOR RR1 response rates by treatment group, and within the experimental design strata (high- and low-likelihood of a Spanish speaker

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

Table 1. Overall Response Rates for English-only and Spanish/English and by Spanish-Speaker Likelihood Stratum, and Sampling Domain

	English-only materials			Spanish/English materials			Chi-square test *p*-value
	n	Respondents	Response rate (%)	*n*	Respondents	Response rate (%)	
Overall	16,278	2,270	14.0	3,103	373	12.0	<.01
Spanish-speaker at address							
High likelihood	1,458	168	11.5	1,456	149	10.2	.29
Low likelihood	14,820	2,102	14.2	1,647	224	13.6	.54
Sampling domain							
<10% Black, <10% Hispanic	5,504	866	15.7	644	91	14.1	.32
>10% Black, <10% Hispanic	4,348	588	13.5	552	73	13.2	.90
<10% Black, >10% Hispanic	3,848	520	13.5	1,283	145	11.3	.05
>10% Black, >10% Hispanic	2,578	296	11.5	624	64	10.3	.42

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

within the HH) and sampling domains, which are based upon the race and ethnicity distribution of the population at the Census block group level. The inclusion of the two-sided English/Spanish letter significantly decreased the overall response rate but did not significantly change the screener response rate among the high-likelihood of Spanish-speaking addresses ($n = 2{,}914$). The two-sided letter also did not significantly change the screener response rate for the 10 percent of cases mailed this letter among the low-likelihood addresses ($n = 16{,}467$). The only domain where there was a significant difference in response rates between the two treatments was the Domain that is <10 percent Black and >10 percent Hispanic ($n = 5{,}131$) with the bilingual materials producing a lower screening response rate.

Table 2 presents information about the race and ethnicity of cases that were selected for the main survey among HHs that completed a screening survey. The results are broken down by experimental treatment and experimental design stratum. Overall, there were no significant differences in the distribution of race/ethnicity across the two treatments. We did find that among high-likelihood Spanish-speaking HHs that *completed the screener* and had eligible persons present ($n = 236$), the percentage of respondents selected for the main survey that was Hispanic was about 6 percentage points higher for HHs receiving the two-sided letter (76 versus 70 percent). This difference was only 11 versus 9 percent for low-likelihood cases that completed the screener ($n = 1{,}274$). These differences are not significant. The predictive modeling, although not optimized directly for this target group, clearly worked well for identifying HHs likely to contain someone of Hispanic descent.

4.2 Experiment 2

Table 3 shows the conditional screening response rates and sample sizes for the two treatment conditions. The use of the $5 incentive along with the priority mailer led to a large and significant increase in response rates relative to no additional mailing or incentive (13.8 versus 5.3 percent, $p < .0001$).

These results remained robust in the multivariable logit model (see table A.2 for details). The adjusted odds ratio for the $5/priority mailer was 2.91 (95 percent CI: 2.60, 3.27, $p < .001$), suggesting a nearly three-fold increase in the odds of responding for those "difficult" sampled HHs receiving the $5/priority mailer. There were no notable interactions between the treatment indicator and other predictors in either the logit model or when using classification trees, suggesting that the treatment worked equally well for cases from all groups that can be defined using the predictors in the model (see table A.2 for details).

There were no differences among screener survey informants with respect to any of the characteristics measured in that short survey. All characteristics are

Table 2. Percentage Hispanic of Eligible Cases Selected for the Main Survey by Treatment and by Likelihood of Spanish Speaker Stratum

	English-only materials		Spanish/ English materials		Chi-square test p-value
	n	%	n	%	
High-likelihood Spanish speaker					.4985
Hispanic	83	70	90	76	
Non-Hispanic Black	3	3	4	3	
Non-Hispanic White	25	21	21	18	
Non-Hispanic Other	7	6	3	3	
Low-likelihood Spanish speaker					.7997
Hispanic	103	9	13	11	
Non-Hispanic Black	171	15	20	17	
Non-Hispanic White	746	65	75	63	
Non-Hispanic Other	134	12	12	10	

Table 3. Response Rates for $5/Priority Mailer versus No Intervention

Treatment	No intervention	$5/priority mailer
n	12,649	4,830
Response rate (%)	5.3	13.8*

*Chi-square test: $p < .0001$.

presented in table 4. As a reminder, the "no intervention" group only includes screening surveys conducted after the intervention was implemented.

Next, we review the characteristics of sampled persons within eligible HHs—that is, among HHs with at least one eligible person, what are the characteristics of the sampled persons? Not all sampled persons will agree to complete the main survey. In this case, we do find that the experimental treatment did appear to appeal to Black persons who were age-eligible, resulting in a higher proportion of Black persons sampled for the main interview. These results are presented in table 5. The "no intervention" group includes eligible, sampled persons identified in the control group after the intervention occurred. The persons who were eligible and sampled within HHs were more likely to be Black under the $5 + priority mailer treatment (19 percent for the experimental group versus 12 percent for the control group). Thus, the intervention increased the design-weighted proportion of eligible and sampled Black

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

Table 4. Demographic Characteristics (Measured from the Screening Survey) of Screening Survey Informants by Treatment

	No intervention ($n = 672$)		$5+ priority mailer ($n = 668$)		Chi-square *p*-value
	Respondents	%	Respondents	%	
Sex					.443
Male	239	37	250	39	
Female	415	63	395	61	
Age					.349
<25	30	5	33	5	
25–34	94	15	81	13	
35–49	142	22	165	26	
50+	382	59	363	57	
Race/ethnicity					.570
Hispanic	79	12	86	13	
Non-Hispanic White	433	66	408	63	
Non-Hispanic Black	96	15	110	17	
Non-Hispanic Other	44	7	42	7	
Number of adults					.145
1	235	35	223	33	
2	292	44	269	40	
3	75	11	100	15	
4+	66	10	75	11	

persons from 7.9 percent prior to the intervention to 10.8 percent after the intervention (i.e., closer to the population proportion of 12.4 percent; Jones, Marks, Ramirez, and Rios-Vargas 2021).

Finally, we look at the characteristics of eligible individuals randomly selected for the main survey who also *responded* to the main survey, to see if this apparent impact carries through to that stage. Table 6 shows the characteristics of the main survey respondents by treatment. These data come from the main survey and, therefore, include characteristics (education and marital status) not available among main survey nonrespondents. For this group, the experimental intervention is associated with significant changes in the race/ethnicity and marginally significant changes in the educational composition of the respondents. The experimental treatment recruited higher proportions of Black and Hispanic respondents. Furthermore, the experimental group seemed to recruit people with lower educational attainment (high school or less and some college). Additional characteristics from the main survey are presented in table A.3.

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

Table 5. Race/Ethnicity (from the Screening Survey) of Eligible, Sampled Persons by Treatment

	No intervention ($n = 324$)		\$5+ priority mailer ($n = 340$)		Chi-square test p-value
	Respondents	%	Respondents	%	
Race/ethnicity					.036
Hispanic	60	19	71	21	
Non-Hispanic White	193	60	170	50	
Non-Hispanic Black	39	12	65	19	
Non-Hispanic Other	32	10	34	10	

5. DISCUSSION

5.1 Summary of Results

With regard to our first experiment, we found that bilingual screener invitation materials (English/Spanish) produced slightly lower overall screening response rates and did not increase screening response rates among addresses predicted to have a Spanish speaker. This runs counter to previous findings (Bouffard and Tancreto 2006; Brick et al. 2012; Elliott et al. 2019). This might be evidence of a "backfire" effect—that is, the bilingual materials might reduce response rates among targeted subgroups in the population. The decrease in response rate is relatively small and might be outweighed by an eventual positive effect of the bilingual materials: among high-likelihood Spanish-speaking HHs that completed the screening questionnaire, the bilingual materials did increase the percentage of respondents selected for the main survey who were Hispanic. Furthermore, 74 percent of the respondents from these high-likelihood HHs selected for the main survey were Hispanic, compared to only 9 percent of the respondents from low-likelihood HHs. Consistent with Elliott et al. (2019), our predictive model was highly effective at identifying HHs containing Hispanic persons.

Our screening protocol for the AFHS sought to identify HHs with eligible individuals between the ages of 18 and 49. The protocol included three phases: an initial mailed invitation with a \$2 incentive to complete the screening questionnaire online, a postcard reminder to complete the screener online, and a mailed reminder including a paper version of the screening questionnaire. Among HHs for which none of these screening invitations was effective, we found in a randomized experiment that a final mailing including an additional \$5 incentive in a priority mailer was highly effective at increasing screener response rates, consistently across all subgroups of interest. This is essentially a replication of the finding by Messer and Dillman (2011), with the additional result that this works well across a variety of

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

Table 6. Characteristics (Measured in the Main Survey) of Main Survey Respondents by Treatment

	No intervention ($n = 205$)		\$5+ priority mailer ($n = 190$)		Chi-square *p*-value
	Respondents	%	Respondents	%	
Sex					.361
Male	75	37	79	42	
Female	130	63	111	58	
Age					.902
18–24	41	20	38	20	
25–29	33	16	26	14	
30–34	45	22	43	23	
35–39	34	17	39	21	
40–44	31	15	28	15	
45–49	21	10	16	8	
Race/ethnicity					.026
Hispanic	29	14	39	21	
Non-Hispanic White	137	67	101	53	
Non-Hispanic Black	19	9	31	16	
Non-Hispanic Other	20	10	19	10	
Education					.053
High school or less	31	15	44	23	
Some college	61	30	66	35	
4-Year college	50	24	33	17	
More than college	63	31	47	25	
Marital status					.791
Married	78	38	63	33	
With partner	26	13	19	10	
Widowed	0	0	1	1	
Divorced	11	5	10	5	
Separated	5	2	5	3	
Never married	85	41	92	48	

subgroups, including Hispanics, non-Hispanic Blacks, those with a high school education or less, and persons who have never been married.

Unfortunately, with this design, we were unable to determine if the increase in screener response rates was due to the \$5 incentive, the priority mailing, or any mailing at all, and disentangle their independent effects.

5.2 Implications for Future Screening Operations and Research

We find that using bilingual invitation letters consistently may serve to increase the percentage of Hispanic individuals who are selected for the main survey.

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

However, in contrast with previous findings, the use of bilingual materials also led to a lower screening response rate. Furthermore, a small additional incentive combined with a different mailing for a *subsample* of cases that have not yet responded to a screening questionnaire increases the number of completed screening surveys and response rates. Furthermore, the approach might be used to raise response rates among low-responding groups. But are both of these design changes—the additional incentive or the priority mailer—necessary to obtain this result? Future experiments need to decompose their independent effects.

Supplementary Materials

Supplementary materials are available online at academic.oup.com/jssam.

REFERENCES

AAPOR (2016), *Standard Dispositions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (9th ed.). https://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx.

Biemer, P. P., J. Murphy, S. Zimmer, C. Berry, G. Deng, and K. Lewis (2018), "Using Bonus Monetary Incentives to Encourage Web Response in Mixed-Mode Household Surveys," *Journal of Survey Statistics and Methodology*, 6, 240–261.

Bouffard, J. A., and J. G. Tancreto (2006), "Experimental Treatment Results of the Bilingual Census Form from the 2005 National Census Test," *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 2779–2786.

Brick, J. M., A. Caporaso, D. Williams, and D. Cantor (2019), "Reaching Spanish-Speakers in Self-Administered Surveys," *Archives of Psychology*, 3, 1–14.

Brick, J. M., J. M. Montaquila, D. Han, and D. Williams (2012), "Improving Response Rates for Spanish Speakers in Two-Phase Mail Surveys," *Public Opinion Quarterly*, 76, 721–732.

Bucks, B.,. M. P. Couper, and S. L. Fulford (2020), "A Mixed-Mode and Incentive Experiment Using Administrative Data," *Journal of Survey Statistics and Methodology*, 8, 352–369.

Church, A. H. (1993), "Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis," *Public Opinion Quarterly*, 57, 62–79.

DeBell, M., M. Amsbary, V. Meldener, S. Brock, and N. Maisel (2018a), "Methodology Report for the ANES 2016 Time Series Study," Technical Report. ANES, Stanford University and the University of Michigan.

DeBell, M., S. Jackman, N. Maisel, M. Amsbary, V. Meldener, J. M. Brick, M. Krupenkin, and E. Peterson (2018b), "Methodology and Findings of the ANES 2016 Recruitment Panel Study," Technical Report. ANES, Stanford University and the University of Michigan.

Dillman, D. A. (1978), *Mail and Telephone Surveys: The Total Design Method*, New York: Wiley.

Dillman, D. A., J. D. Smyth, and L. M. Christian (2014), *Internet, Phone, Mail, and Mixed-Mode Surveys the Tailored Design Method*, Hoboken: Wiley.

Dirksz, G., P. J. Lavrakas, L. Lusskin, B. Ponce, and P. Felstead (2017), "Testing the Effects of Token Noncontingent Incentives in Follow-Up Mailings in a Mixed Mode Survey," paper presented at the Annual Conference of the American Association for Public Opinion Research, New Orleans, LA, May.

Dykema, J., K. Jaques, K. Cyffka, N. Assad, R. G. Hammers, K. Elver, K. C. Malecki, and J. Stevenson (2015), "Effects of Sequential Prepaid Incentives and Envelope Messaging in Mail Surveys," *Public Opinion Quarterly*, 79, 906–931.

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022

Edwards, P., I. Roberts, M. Clarke, C. DiGuiseppi, S. Pratap, R. Wentz, and I. Kwan (2002), "Increasing Response Rates to Postal Questionnaires: Systematic Review," *British Medical Journal*, 324, 1183.

Elliott, M. N., D. J. Klein, P. Kallaur, J. A. Brown, R. D. Hays, N. Orr, A. M. Zaslavsky, M. K. Beckett, S. Gaillot, C. A. Edwards, and A. M. Haviland (2019), "Using Predicted Spanish Preference to Target Bilingual Mailings in a Mail Survey with Telephone Follow-Up," *Health Services Research*, 54, 5–12.

Groves, R. M., R. B. Cialdini, and M. P. Couper (1992), "Understanding the Decision to Participate in a Survey," *Public Opinion Quarterly*, 56, 475–495.

Groves, R. M., E. Singer, and A. Corning (2000), "Leverage-Saliency Theory of Survey Participation: Description and an Illustration," *Public Opinion Quarterly*, 64, 299–308.

Jones, N., R. Marks, R. Ramirez, and M. Rios-Vargas (2021), "2020 Census Illuminates Racial and Ethnic Composition of the Country," US Census Bureau website. Available at https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html.

Koslow, S., P. N. Shamdasani, and E. E. Touchstone (1994), "Exploring Language Effects in Ethnic Advertising: A Sociolinguistic Perspective," *Journal of Consumer Research*, 20, 575–585.

Kropf, M. E., and J. Blair (2005), "Eliciting Survey Cooperation: Incentives, Self-Interest, and Norms of Cooperation," *Evaluation Review*, 29, 559–575.

Mercer, A., A. Caporaso, D. Cantor, and R. Townsend (2015), "How Much Gets You How Much?: Monetary Incentives and Response Rates in Household Surveys," *Public Opinion Quarterly*, 79, 105–129.

Messer, B. L., and D. A. Dillman (2011), "Surveying the General Public over the Internet Using Address-Based Sampling and Mail Contact Procedures," *Public Opinion Quarterly*, 75, 429–457.

Wagner, J., J. Arrieta, H. Guyer, and M. B. Ofstedal (2014), "Does Sequence Matter in Multimode Surveys: Results from an Experiment," *Field Methods*, 26, 141–155.PMCID: PMC3992480

Williams, D, and., J. M. Brick (2018). "Trends in U.S. Face-to-Face Household Survey Nonresponse and Level of Effort," *Journal of Survey Statistics and Methodology*, 6, 186–211.

Downloaded from https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smac024/6640340 by University of Michigan user on 14 July 2022