PLOS ONE

Methods for Improving the Quality of National Household Surveys

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Article
Full Title:	Methods for Improving the Quality of National Household Surveys
Short Title:	Methods for Improving the Quality of National Household Surveys
Corresponding Author:	Brady T. West University of Michigan Ann Arbor, MI UNITED STATES
Keywords:	survey methodology; mixed-mode survey design; web surveys; survey nonresponse
Abstract:	Increasing access to the internet, the increasing costs of large-scale face-to-face data collections, and the general reluctance of the public to participate in intrusive in-person data collections all mean that new approaches to nationally representative surveys are urgently needed. The COVID-19 pandemic accelerated the need for faster, higher-quality alternatives to face-to-face data collection. These trends place a high priority on the evaluation of innovative web-based data collection methods that are convenient for the public and yield scientific information of high quality. The web mode is particularly appealing because it is relatively inexpensive, it is logistically flexible to implement, and it affords a high level of privacy and confidentiality when correctly implemented. With this study, we aimed to conduct a methodological evaluation of a sequential mixed-mode web/mail data collection protocol, including modular survey design concepts, which was implemented on a national probability sample in the U.S. in 2020-2021. We implemented a series of experiments to test theoretically-informed hypotheses that 1) the use of mail and increased incentives to follow up with households that did not respond to an invitation to complete a household screening questionnaire online would help to recruit different types of households; 2) the use of mail and telephone reminders to follow up with individuals invited to complete the main survey but not yet responding would further improve the quality of the respondent sample; and 3) the use of modular survey design, which involves splitting a lengthy self-administered survey up into multiple parts that can be completed at a respondent's convenience, would improve survey completion rates. We find support for the first two hypotheses, suggesting a robust new approach for conducting time-saving, cost-efficient web/mail data collections with national samples that also have increased statistical efficiency.
Order of Authors:	Brady T. West
	Shiyu Zhang
	James Wagner
	Rebecca Gatward
	Htay-Wah Saw
	William G. Axinn
Opposed Reviewers:	
Additional Information:	
Question	**Response**
Financial Disclosure Enter a financial disclosure statement that describes the sources of funding for the work included in this submission. Review the submission guidelines for detailed requirements. View published research	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; PI: B.T. West; website: https://www.nichd.nih.gov/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

articles from *PLOS ONE* for specific examples. This statement is required for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate. **Unfunded studies** Enter: *The author(s) received no specific funding for this work.* **Funded studies** Enter a statement with the following details: • Initials of the authors who received each award • Grant numbers awarded to each author • The full name of each funder • URL of each funder website • Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript? • **NO** - Include this sentence at the end of your statement: *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.* • **YES** - Specify the role(s) played. * typeset	
Competing Interests Use the instructions below to enter a competing interest statement for this submission. On behalf of all authors, disclose any competing interests that could be perceived to bias this work—acknowledging all financial support and any other relevant financial or non-financial competing interests. This statement is required for submission and **will appear in the published article** if the submission is accepted. Please make sure it is accurate and that any funding sources listed in your Funding Information later in the submission form are also declared in your Financial Disclosure statement.	The authors have declared that no competing interests exist.

View published research articles from *PLOS ONE* for specific examples. **NO authors have competing interests** Enter: *The authors have declared that no competing interests exist.* **Authors with competing interests** Enter competing interest details beginning with this statement: *I have read the journal's policy and the authors of this manuscript have the following competing interests: [insert competing interests here]* * typeset	
Ethics Statement Enter an ethics statement for this submission. This statement is required if the study involved: • Human participants • Human specimens or tissue • Vertebrate animals or cephalopods • Vertebrate embryos or tissues • Field research Write "N/A" if the submission does not require an ethics statement. General guidance is provided below. Consult the submission guidelines for detailed instructions. **Make sure that all information entered here is included in the Methods section of the manuscript.**	All study procedures, materials, and methods were approved by the Human Subjects Institutional Review Board at the University of Michigan (Study HUM00167171). A consent form describing the minimal risks and benefits of the main study, the study procedures, and an assurance of confidentiality was presented either electronically (via the web) or on paper (for mail respondents) prior to the main survey instrument. Web respondents were asked to agree to the terms described before proceeding.

Format for specific study types **Human Subject Research (involving human participants and/or tissue)** • Give the name of the institutional review board or ethics committee that approved the study • Include the approval number and/or a statement indicating approval of this research • Indicate the form of consent obtained (written/oral) or the reason that consent was not obtained (e.g. the data were analyzed anonymously) **Animal Research (involving vertebrate animals, embryos or tissues)** • Provide the name of the Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board that reviewed the study protocol, and indicate whether they approved this research or granted a formal waiver of ethical approval • Include an approval number if one was obtained • If the study involved *non-human primates*, add *additional details* about animal welfare and steps taken to ameliorate suffering • If anesthesia, euthanasia, or any kind of animal sacrifice is part of the study, include briefly which substances and/or methods were applied **Field Research** Include the following details if this study involves the collection of plant, animal, or other materials from a natural setting: • Field permit number • Name of the institution or relevant body that granted permission	
Data Availability Authors are required to make all data underlying the findings described fully available, without restriction, and from the time of publication. PLOS allows rare exceptions to address legal and ethical concerns. See the PLOS Data Policy and FAQ for detailed information.	No - some restrictions will apply

A Data Availability Statement describing where the data can be found is required at submission. Your answers to this question constitute the Data Availability Statement and **will be published in the article**, if accepted. **Important:** Stating 'data available on request from the author' is not sufficient. If your data are only available upon request, select 'No' for the first question and explain your exceptional situation in the text box. Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?	
Describe where the data may be found in full sentences. If you are copying our sample text, replace any instances of XXX with the appropriate details. • If the data are **held or will be held in a public repository**, include URLs, accession numbers or DOIs. If this information will only be available after acceptance, indicate this by ticking the box below. For example: *All XXX files are available from the XXX database (accession number(s) XXX, XXX.*). • If the data are all contained **within the manuscript and/or Supporting Information files**, enter the following: *All relevant data are within the manuscript and its Supporting Information files.* • If neither of these applies but you are able to provide **details of access elsewhere**, with or without limitations, please do so. For example: *Data cannot be shared publicly because of [XXX]. Data are available from the XXX Institutional Data Access / Ethics Committee (contact via XXX) for researchers who meet the criteria for access to confidential data.* *The data underlying the results presented in the study are available from (include the name of the third party*	At the time of this submission, we are working with the Inter-university Consortium for Political and Social Research (ICPSR) to make our data publicly available; this process is ongoing, and readers will be able to download our data from the ICPSR website when it is complete. In the meantime, readers can contact Brady T. West (PI of the AFHS; bwest@umich.edu) for access to a working version of the respondent data. To be clear, these data will only represent survey respondents. This methodological article includes analyses of paradata from nonrespondents to our survey, who will not be included in the ICPSR release. Readers can also contact Brady T. West for de-identified information on AFHS nonrespondents that was used in these analyses.

and contact information or URL). • This text is appropriate if the data are owned by a third party and authors do not have permission to share the data. * typeset	
Additional data availability information:	

THE UNIVERSITY OF MICHIGAN
Survey Methodology Program • Survey Research Center • Institute for Social Research
426 Thompson Street, Room 4050, Ann Arbor, MI 48106 • Campus P.O. Box 1248
Phone: (734) 647-3545 • Fax: (734) 764-8263

October 24, 2022

To the Editors of PLOS ONE:

On behalf of my five co-authors, I am submitting this manuscript entitled "Methods for Improving the Quality of National Household Surveys" to be considered for publication as a research article in *PLOS ONE*. We have not had any prior interactions with PLOS regarding this manuscript.

Increasing access to the internet, the increasing costs of large-scale face-to-face data collections, and the general reluctance of the public to participate in intrusive in-person data collections all mean that new approaches to nationally representative surveys are urgently needed. We conducted a methodological evaluation of a sequential mixed-mode web/mail data collection protocol, including modular survey design concepts, which was implemented on a national probability sample in the U.S. in 2020-2021. We implemented a series of experiments to test theoretically-informed hypotheses that 1) the use of mail and increased incentives to follow up with households that did not respond to an invitation to complete a household screening questionnaire online would help to recruit different types of households; 2) the use of mail and telephone reminders to follow up with individuals invited to complete the main survey but not yet responding would further improve the quality of the respondent sample; and 3) the use of modular survey design, which involves splitting a lengthy self-administered survey up into multiple parts that can be completed at a respondent's convenience, would improve survey completion rates. We found strong support for the first two hypotheses, suggesting a robust new approach for conducting time-saving, cost-efficient web/mail data collections with national samples that also have increased statistical efficiency.

We believe that the following five Academic Editors would be best-suited to handle this manuscript: Philip Anglewicz, Wolfgang Himmel, Diwakar Mohan, Candace Nelson, and Amit Patel. We do not have any reviewers in mind that we would oppose.

We look forward to hearing back from you with regard to your thoughts on this manuscript, and whether *PLOS ONE* would be willing to publish this study following peer review.

Best Regards,

Brady West

Brady T. West
Research Professor
Institute for Social Research
University of Michigan-Ann Arbor
bwest@umich.edu

Title:

Methods for Improving the Quality of National Household Surveys

Authors:

Brady T. West

Shiyu Zhang

James Wagner

Rebecca Gatward

Htay-Wah Saw

William G. Axinn

Affiliations:

All authors are affiliated with the Survey Research Center, Institute for Social Research, University of Michigan-Ann Arbor, Ann Arbor, Michigan, USA.

Abstract

Increasing access to the internet, the increasing costs of large-scale face-to-face data collections, and the general reluctance of the public to participate in intrusive in-person data collections all mean that new approaches to nationally representative surveys are urgently needed. The COVID-19 pandemic accelerated the need for faster, higher-quality alternatives to face-to-face data collection. These trends place a high priority on the evaluation of innovative web-based data collection methods that are convenient for the public and yield scientific information of high quality. The web mode is particularly appealing because it is relatively inexpensive, it is logistically flexible to implement, and it affords a high level of privacy and confidentiality when correctly implemented. With this study, we aimed to conduct a methodological evaluation of a sequential mixed-mode web/mail data collection protocol, including modular survey design concepts, which was implemented on a national probability sample in the U.S. in 2020-2021. We implemented a series of experiments to test theoretically-informed hypotheses that 1) the use of mail and increased incentives to follow up with households that did not respond to an invitation to complete a household screening questionnaire online would help to recruit different types of households; 2) the use of mail and telephone reminders to follow up with individuals invited to complete the main survey but not yet responding would further improve the quality of the respondent sample; and 3) the use of modular survey design, which involves splitting a lengthy self-administered survey up into multiple parts that can be completed at a respondent's convenience, would improve survey completion rates. We find support for the first two hypotheses, suggesting a robust new approach for conducting time-saving, cost-efficient web/mail data collections with national samples that also have increased statistical efficiency.

Introduction

As internet access has continued to spread worldwide, conducting surveys via the web has become an attractive option for scientific research on the general population to advance social science and policy evaluation. The primary advantages of web surveys include portability, flexibility, and confidentiality – web surveys allow respondents to complete surveys at whatever time and location is convenient and private for them. Collecting survey data via the web is also an attractive method for lowering data collection costs and curtailing data collection timeframes, both of which support greater scientific innovation [1-3]. These advantages extend when respondents are allowed to use multiple devices, including personal computers, laptops, tablets, and smartphones, further providing respondents with more options for convenience with little difference in measurement error between the devices [4-7]. This type of data collection approach is particularly attractive given the COVID-19 pandemic, which has strained face-to-face data collection operations globally and increased the need for the collection of population health data in a timely manner.

On the other hand, the advantages of the web mode can be offset by the serious disadvantages of non-universal access to the internet (and the coverage bias that it may introduce in survey research [8]), lower response rates compared to non-web modes [2, 9, 10], and the potential for nonresponse bias to mislead investigators. In practice, these issues can be remedied with sequential mixed-mode designs, where alternative non-web modes such as mail and telephone are used to follow up with non-respondents. This approach is informed by the leverage-salience theory of survey participation [11], where exclusive use of a single mode of data collection may not appeal to all individuals selected for a sample, and the use of alternative, complementary data

collection protocols for nonresponse follow-up may help to recruit different types of respondents for whom the initial protocol was not appealing [12]. Sequential mixed-mode approaches have been shown to increase response rates while also possibly decreasing the nonresponse bias in survey estimates compared to single-mode surveys [1, 13-15].

Guiding Theoretical Framework

Fortunately, while the general population's access to the internet has improved, the survey methodology of augmenting web surveys to provide a more robust representation of the general population has also advanced. Here we harness the key hypotheses from those advances to compare their ability to improve the population representation in surveys primarily implemented using the internet. First, we consider the initial household screening required to implement targeted (for example by age eligibility) surveys of individuals, building on what has been learned about the benefits of sequential mixed-mode designs and implementation of cash incentives. Second, we investigate the specific benefits of targeted reminder contacts with individuals who have not responded. Third, we investigate the potential of modular designs that aim to lower the burden of participating in surveys and spread the participation effort across time to improve representation. Our explicit, simultaneous evaluation of this full set of options derived from recent advances in survey methodology allows us to reach a robust set of recommendations for new approaches to future surveys.

The Use of Sequential Mixed-Mode Approaches and Cash Incentives for Household Screening

Despite the clear potential of these sequential mixed-mode approaches, there is insufficient research evaluating the utility of using self-administered modes of data collection (e.g., web and

mail) to measure complex, cross-sectional, national probability samples, where separate household screening and main survey data collection stages are necessary. Many national studies are beginning to transition to such a data collection approach, largely out of a need to reduce costs; see [13] for details. Household screening is often needed in such data collections to determine whether there are persons present in a sampled household who are eligible for a particular survey, where one or more eligible individuals in a household are then randomly selected and invited to complete the main survey by either web or paper. Very little existing literature has evaluated effective methods for increasing household response rates when sampled households are invited to complete a screening questionnaire online. Those studies find general support for protocols grounded in both leverage-salience theory and social exchange theory [16]. In the latter case, the provision of incentives to sampled households or persons who are not intrinsically motivated by the topic of the survey may introduce extrinsic motivation to participate, along with feelings of a need to reciprocate if provided with a cash incentive by the data collector.

For example, the 2016 American National Election Study (ANES) included an internet sample in addition to a “standard” face-to-face sample. For the internet portion of the study, mail invitations including cash incentives (initially $10 or $20 prepaid as part of an experiment, and eventually escalating to promised offers of $40 and $80 for completion of a pre-election and post-election interview) were sent to a probability sample of addresses [17, 18]. Sampled households were initially invited to complete a brief screening questionnaire online, after which an adult in the household was randomly selected. If the household informant (i.e., the person supplying the information needed to randomly select one household member) was not selected,

he or she was asked for additional information about the household and contact information for the selected respondent. If the informant was randomly selected, then the main survey started. The screening response rate was about 57%, which rivals the response rates seen today by many face-to-face surveys [19, 20]. See [17] for additional details.

Another study focused on a national survey of parents about the health of their children (the National Survey of Children's Health, or NSCH; see http://childhealthdata.org/learn/NSCH for details). This survey has been conducted using web/mail since 2016 by the U.S. Census Bureau. In 2016, the median survey length was 4.2 minutes for the screening interview and 26.5 minutes for the main survey for those with children; these times were much smaller for those without children. The screener completion rate was 53.0%, and the weighted completion rate for the main survey was 33.0%, yielding a weighted overall response rate of 40.7%. The 2016 study also included a screener incentive experiment ($0, $2, $5). Screener response rates were 50.3%, 53.2% and 55.3% respectively, finding support for the use of higher incentives at the screening stage. They also experimentally evaluated the use of an additional cash incentive ($0, $2, $5, $10) in the third mailing for the main survey, finding that the use of additional (i.e., non-zero) incentives in nonresponse follow-up helped to increase the main response rates further.

The National Household Education Surveys (NHES) obtained a 69% response rate for their screening survey in a large 2011 field test that involved converting from a telephone mode of data collection to mail (see [21-23] for details about this conversion, which improved coverage of the target population and increased the response rate for essentially the same cost). They tested $2 versus $5 incentives and found that the $2 incentive for the screener obtained a 67%

response rate, and the $5 incentive did not increase the response rate significantly [24]. This study suggested that a $2 pre-paid incentive may be sufficient for invoking social exchange theory and improving response rates for self-administered screening questionnaires.

Prior to 2015, the Residential Energy Consumption Survey (RECS) was conducted entirely as a national face-to-face survey. In 2015, RECS implemented a web/mail survey in parallel with their face-to-face survey. As the RECS concerns features of the housing unit and not individuals, there is no screening operation to identify eligible persons, but we can still draw on their research to understand household decisions to complete self-administered questionnaires at the screening stage of a data collection. The implementation of the web/mail approach indicated that offering 1) the choice of web or mail completion, 2) a higher incentive (a $10 bonus) for web completion, and 3) mail follow-up using a brief, two-page questionnaire for nonrespondents, produced response rates between 54% and 56%, while only slightly increasing costs and maintaining the consistency of sample characteristics with American Community Survey benchmarks for the target population (despite lower response rates compared to the face-to-face data collection; [25-27]. The use of different protocols at different times provides motivation for different types of participants, as suggested by leverage-salience theory. The 2020 RECS is now running exclusively as a self-administered web/mail survey because of these earlier studies.

Revisiting the NHES studies discussed earlier, they also found that a modest $5 incentive resulted in a sample that was more closely aligned with population benchmarks while also maintaining a response rate at the main stage that exceeded 70% [24]. This provides additional

support for the ability of modest incentives to secure more balanced participation from different population subgroups in self-administered surveys.

Collectively, we therefore have both theoretical and empirical support for a first hypothesis that *a sequential mixed-mode approach to household screening, offering mail options and additional cash incentives to persons who do not respond initially to a web invitation provided via a mailed letter, will improve rates of response to a self-administered household screening questionnaire, along with the quality of the responding household sample.*

Reminders for Individuals Selected for the Main Survey

Once a household completes a self-administered screening questionnaire and the data collector invites one or more randomly selected eligible persons from the household to complete the main study using a self-administered mode, additional effort may be needed to motivate respondents to complete the primary survey of interest. While we know that these individuals are already somewhat motivated, either intrinsically or extrinsically, given that their households completed the screening questionnaire, they are still being asked to complete a second, longer survey via the web or a paper questionnaire. Additional motivation of the same types described earlier may prove effective, including the use of different modes that have not been attempted before to remind these individuals about the benefits of participating in the main study [12]. For example, the NHES found that varying the method of contact (including use of priority mailings) and following up nonrespondents with telephone prompts/reminders tended to work well for increasing response rates to their main survey [28].

These results give rise to a second hypothesis that *the use of mail and telephone reminders for motivated participants whose households have already completed the screening questionnaire will improve rates of response to the main survey request*. What remains unknown at the main stage of data collection is whether these varied reminders will also serve to improve the quality of the resulting respondent sample, considering the characteristics of responding individuals. We aim to examine this question with our analysis.

Modular Design for the Main Survey

Finally, there is a need to make the survey experience as convenient as possible for persons agreeing to complete the main self-administered survey. Lengthy web surveys may lead to survey breakoff [29, 30], where individuals start the main survey online but fail to complete it. Recent studies have suggested that respondents are not willing to spend lengthy amounts of time on web surveys, and especially those completed on mobile devices [31], making survey length and the burden of the response task very important considerations for designers of web surveys. *Modular survey design*, where a survey researcher explicitly divides a survey up into shorter parts (or "chunks") that a respondent can complete at different points in time, has been the subject of several recent methodological studies [31-37]. In theory, modular design would be expected to increase survey completion rates by reducing the *perceived burden* of a lengthy survey response task [32, 34, 35]. Innovative design strategies combined with the modular approach may also increase completion rates: for example, if participants are notified in advance that they will only need to answer a small number of questions at different points in time, or that they will receive a micro-incentive each time that they respond to a small set of questions

(drawing on the cost-benefit theory of survey response; [16]), completion rates may increase overall [38, 39].

The emerging literature in this area has presented some evidence of positive benefits associated with the modular design approach. Initial experimental studies have suggested increased respondent convenience, increased initial rates of participation, and higher data quality among participating individuals [32, 34, 35]. Non-experimental studies in various international settings have generally echoed these findings [38, 40, 41]. However, other recent studies of this approach have produced mixed or negative findings, specifically with respect to overall survey *completion* rates and item nonresponse rates [32, 33, 35, 42, 43]. These studies have consistently called for more research into the lengths of survey modules and the appropriate amount of time between the modules.

Another recent national study that attempted to transition a general social survey (designed for a cross-sectional national sample) from the face-to-face mode to a self-administered web/mail mode is the German component of the 2017/2018 European Values Study (EVS). [44] describe the results of a three-arm experiment demonstrating that a concurrent web/mail approach and matrix sampling (i.e., assigning random subsets of non-core questionnaire items to different subsamples; [45-48]) resulted in a viable and more cost-efficient alternative to face-to-face data collection, yielding shorter median interview lengths and fielding times (38 minute median interview length for the survey over 6-8 weeks, versus 59 minutes for the survey over 6 months) and higher response rates (36.1% vs. 28.0%) relative to the face-to-face approach. One outstanding issue based on the experiment was slightly worse representation of the general

population by self-administered respondents, with the face-to-face approach yielding more younger and non-German respondents (among other slight differences relative to benchmarks). Positive results were also reported for similar transitions of the EVS in Denmark, Iceland, and Switzerland by [49].

We therefore see evidence in the literature of efforts to reduce survey length and respondent burden (via modular design or matrix sampling) being beneficial for participation in self-administered surveys. For additional evidence of the benefits of reducing survey length for participation in web and mail surveys, see [50-53]. To our knowledge, no studies have evaluated modular design as a tool for increasing survey completion rates when a sequential mixed-mode web/mail approach is used to measure a national sample. Given the evidence above, we present a third hypothesis that *a carefully designed and communicated modular survey protocol will increase completion rates for a lengthy main web/mail survey due to an increase in the perceived convenience of participating in the survey*. Although the literature is more mixed in this area, almost nothing is known about whether the effectiveness of the modular design approach varies across socio-demographic subgroups in this web/mail setting, which has important implications for future adaptive survey design strategies in this context. We seek to examine this question as part of our analysis as well.

Materials and Methods

Overview of the American Family Health Study

In April 2020, a new project known as the American Family Health Study (AFHS; see afhs.isr.umich.edu) initiated data collection with a national address-based probability sample of

more than 19,000 U.S. addresses. The AFHS used a sequential mixed-mode mail/web protocol for push-to-web household (HH) screening to identify eligible persons aged 18-49. Selected eligible persons were then invited to complete a 60-minute web survey on reproductive health and family formation topics, using a second sequential mixed-mode mail/web protocol that encouraged the selected persons to respond to the "main" survey via the web. This design enabled testing of the first two hypotheses communicated above.

The questionnaire was translated from a face-to-face national survey: The National Survey of Family Growth (NSFG). A consent form describing the minimal risks and benefits of the main study, the study procedures, and an assurance of confidentiality was presented either electronically (via the web) or on paper (for mail respondents) prior to the main survey instrument. Web respondents were asked to agree to the terms described before proceeding. Data collection continued until June 2021.

About half of eligible respondents selected from sampled households that completed the screening questionnaire were randomly assigned to the aforementioned "modular" condition, where the main survey data collection would involve completing three 20-minute surveys allowing for a two-week break before the next survey invitation. The remaining half was assigned to a "full" survey condition, where they were asked to complete the full 60-minute survey in a single sitting. Respondents in the full condition were able to take breaks and return to the survey where they left off. This aspect of the design enabled testing of the third hypothesis communicated above. Additional details regarding the AFHS sample design and data collection

methodology can be found elsewhere (see https://afhs.isr.umich.edu/about-the-study/afhs-methodology/; [54]).

AFHS Screening Protocol

The AFHS screening questionnaire was designed to collect a list of persons aged 18 years and over in the household. In the first phase of this protocol, we selected a stratified probability sample of addresses, oversampling addresses predicted to have an age-eligible (18-49 years old) person present and located in high-density minority areas. Sampled households received a mailed invitation (including a $2 cash incentive) addressed to the resident of a particular state, inviting an adult member of the household to complete a screening questionnaire online. 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.

In the third phase of screening, a follow-up mailing that included a paper version of the screening questionnaire was sent two weeks after the postcard. In the fourth phase of screening, 28 days after the initial invitation, a random subsample of 5,000 non-finalized sampled addresses was sent a priority mailing with a final invitation to complete the screening questionnaire and an additional $5 incentive. Information obtained from completed screening questionnaires was used to identify eligible persons within the sampled households. If there was only one eligible person in the household, then that person was immediately invited to complete the main AFHS survey online. If there was more than one eligible person, one person was randomly selected and then invited to complete the main survey.

AFHS Main Data Collection Protocol

Once an eligible respondent was randomly selected from a sampled household 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. For the screener respondents who responded via web and were selected for the main interview, the initial invitation immediately followed the screening interview (i.e., people were redirected to another web page starting the main interview). For all other cases, the initial invitation was scheduled 14 days after the screener interview was completed. 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 in the screening questionnaire). For eligible nonrespondents for whom we did not have an email or text-enabled phone number, we mailed a follow-up letter 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% of these nonrespondents had telephone numbers available – although some (12%) 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.

This main data collection protocol was repeated for *each survey module* for those selected respondents randomly assigned to the modular condition. These respondents were promised $20 for completing the first module, $20 for completing the second module, and $30 for completing the third module. Figure 1 summarizes the data collection protocols for the screening and main stages of the data collection.

<< INSERT FIGURE 1 HERE >>

Fig. 1: Screening and Main Data Collection Protocols for the AFHS.

All study procedures, materials, and methods were approved by the Human Subjects Institutional Review Board at the University of Michigan (Study HUM00167171).

Data Analysis

To test our first hypothesis (*a sequential mixed-mode approach to household screening, offering mail options and additional cash incentives to persons who do not respond initially to a web invitation provided via a mailed letter, will improve rates of response to a self-administered household screening questionnaire, along with the quality of the responding household sample*), we first evaluated changes in the screener response rate across the four phases of the screening protocol. We considered the phase-specific response rates, cumulative response rates across the four phases, and response rates for each level of the $5 + priority experiment in the fourth phase. We also considered *hypothetical* cumulative response rates, assuming that all nonrespondents either did or did not receive the $5 + priority mailing follow-up in the fourth phase. We then

compared mail and web respondents at each of the screening phases in terms of socio-demographic information collected from the screening questionnaire, to see if the mail follow-up was increasing the diversity of the responding sample.

Next, we used Pearson chi-squared tests to evaluate the effects of the $5 + priority mailing follow-up at the fourth phase of the screening protocol on response rates among active nonrespondents, separately for each of 14 *LifeMode* groups of addresses defined by the *Esri Tapestry Segmentation* (https://www.esri.com/en-us/arcgis/products/data/data-portfolio/tapestry-segmentation) [55]. This system divides U.S. block groups into 14 distinct types of neighborhoods called *LifeModes* based on socio-demographic and socio-economic characteristics; similar commercial data describing different areas can be purchased and linked to sampling frames in other countries [56]. Each of the types is described psycho-graphically using terms like *Affluent Estates*. Although designed for marketing purposes, the Esri segments for the block groups, which can be linked to our full probability sample, have been found to be related to participation rates and mode choices in the 2015 Census Test [57]. See Appendix I of the supplemental materials for more details about this segmentation system, including analyses of the associations of the *LifeModes* with selected key variables in the AFHS. In these comparisons, we also considered the mode of response (web or mail) in each of the experimental groups, to see if the $5 + priority mailing follow-up ultimately produced more web or mail (paper) responses to the screening questionnaire and whether these differences varied across areas defined by the 14 *LifeMode* groups.

To test our second hypothesis (*the use of mail and telephone reminders for motivated participants whose households have already completed the screening questionnaire will improve rates of response to the main survey request*), we first evaluated whether the use of a shortened follow-up mail questionnaire in the third phase of the main protocol improved the diversity of the respondent pool in terms of gender, age, education, and race/ethnicity, separately for the full and modular conditions. When evaluating this diversity, we used estimated distributions from the 2017-2019 NSFG as the population benchmarks. Next, we assessed whether the use of a telephone follow-up reminder at the final phase of the main data collection protocol served to increase response rates for different socio-demographic subgroups. We conducted these analyses separately for each of the experimental conditions (full and modular) to examine whether telephone reminders were more effective when using one of these approaches.

To test our third hypothesis (*a carefully designed and communicated modular survey protocol will increase completion rates for a lengthy main web/mail survey due to an increase in the perceived convenience of participating in the survey*), we compared overall *completion rates* between the full and modular conditions at the main stage of data collection, where a selected person assigned to the modular condition needed to complete all three modules of the survey to be considered as a complete respondent. We also compared module-specific response rates to the overall response rate in the "full" condition. We then performed similar comparisons of the full and modular conditions for different subgroups defined by socio-demographic characteristics collected in the screening questionnaire, to examine whether the modular approach tended to appeal to specific socio-demographic subgroups.

Results

AFHS Response Rates

We obtained an overall response rate in the screening stage of 15.0% and a conditional AAPOR RR4 response rate of 66.0% in the main stage (where for individuals randomly assigned to the modular condition, completing at least two sections of the questionnaire in the first 20-minute module was counted as a partial response). These two rates resulted in a net AAPOR RR4 response rate of 9.9%; we revisit this result in the Discussion. See Appendix II of the supplemental materials for detailed descriptions of these response rate calculations.

Hypothesis 1: Sequential Mixed-Mode Design and Cash Incentives for Nonrespondents Will Improve Screening Outcomes.

Figure 2 presents the sample sizes for each phase (n), obtained by subtracting the counts of respondents and nonrespondents from the available sample in the previous phase. For example, phase 1 had 537 respondents and 5 "hard" nonrespondents (who either called the main study telephone number or wrote a letter back to the study team with a hard refusal), and thus the sample size for phase 2 was equal to 18,839 (i.e., 19,381 – 537 – 5). Figure 2 also indicates the counts of respondents in each phase (n_r), and the corresponding raw phase-specific response rates across the phases of the AFHS screening protocol.

<< INSERT FIGURE 2 HERE >>

Fig. 2: Phase-specific Response Rates at the Screening Stage.

Figure 2 shows that the $5 + priority mailing follow-up in Phase 4 was highly effective in recruiting respondents (see [54] for more details). The response rate in Phase 4a is 8.32 percentage points (p.p.) higher than Phase 4b (where non-respondents did not get an additional $5 + priority follow-up).

Figure 3 presents a *cumulative* assessment of these results. Here, n indicates the overall sample size. The number 17,490 (in Phase 4a and 4b) indicates the number of sample cases entering the fourth phase; these cases were neither respondents nor "hard" nonrespondents in the previous three phases (corresponding to Figure 2 above, 17,490 = 4,951 + 12,539). For the fourth phase where the $5 + priority mailing experiment was conducted, we formed hypothetical scenarios where *all* Phase 4 cases either received (Phase 4a) or did not receive (Phase 4b) the $5 + priority mailing. Because these (hypothetical) response rates have the same denominator (i.e., the full sample), they can be directly subtracted from each other. Thus, the estimated effect of mail packet 1 is 5.78 p.p. (= 11.51% - 5.73%), and the estimated effect of the $5 + priority mailing is an additional 7.51 p.p. (= 19.02% - 11.51%). These results provide general support for our first hypothesis.

<< INSERT FIGURE 3 HERE >>

Fig. 3: Cumulative phase-specific response rates at the screening stage, in addition to hypothetical response rates at the conclusion of Phase 4 if all cases had either received or not received the $5 + priority mailing.

Roughly a third (31%) of screening respondents participated by mailing back a completed paper questionnaire. When comparing the socio-demographic features of the mail and web screener

respondents, we found that mail respondents were more likely than web respondents to be female (64% vs. 58%, $p < 0.01$), older (65% vs. 33% 55+, $p < 0.01$), and Black (17% vs. 11%, $p < 0.01$). We therefore have evidence that the use of a mail (paper) questionnaire for nonresponse follow-up did serve to attract unique types of respondents and increase the diversity of our respondent pool, which is consistent with the prior literature in this area [58].

We further examined whether the $5 + priority mailing differentially motivated cases of different characteristics, and we do so by comparing the phase-4 response rates of sample subgroups defined by Esri Tapestry *LifeMode* groups. In general, we found that the effect of the $5 + priority mailing was *heterogeneous* across *LifeMode* groups. The differences in response rates between Phase 4a (experimental: $5 + priority mailing) and Phase 4b (control: no follow-up) by *LifeMode* and by mode of response are summarized in Table 1. First, this mailing did not significantly increase the response rate for the two *LifeMode* groups *Uptown Individuals* (3.26 p.p. difference between Phase 4a and 4b, $p = 0.208$) and *Scholars and Patriots* (5.17 p.p. difference, $p = 0.160$) (see Appendix I of the supplemental materials for detailed descriptions of these groups). Second, this mailing had the strongest effect in the *LifeMode* group *Rustic Outposts*, increasing its response rate by 11.89 p.p. ($p < 0.0001$). The results suggest that withholding the final mailing from cases in the *Uptown Individuals* and *Scholars and Patriots* groups may be one way to save on mailing costs.

Table 1. Percentage point differences in screening response rates between Phase 4a and 4b by Esri *LifeMode* groups and by mode.

	Phase 4a			Phase 4b			Difference: Phase 4a – Phase 4b					
	Total	by Web	by PAPI	Total	by Web	by PAPI	Total		by Web		by PAPI	
Affluent Estates	13.32	5.43	7.88	5.60	2.70	2.90	7.72	***	2.73	*	4.98	**

Upscale Avenues	12.99	3.94	9.06	5.15	3.03	2.12	7.84	***	0.91	NS	6.94	***
Uptown Individuals	10.45	8.46	1.99	7.19	4.02	3.17	3.26	NS	4.44	*	-1.18	NS
Family Landscapes	14.50	6.39	8.11	4.80	2.40	2.40	9.70	***	3.99	**	5.71	***
GenXurban	16.81	4.37	12.45	7.77	3.01	4.76	9.04	***	1.36	NS	7.69	***
Cozy Country Living	15.08	2.01	13.07	5.36	1.88	3.47	9.72	***	0.13	NS	9.60	**
Ethnic Enclave	9.66	2.62	7.04	4.00	2.08	1.92	5.66	***	0.54	NS	5.12	***
Middle Ground	14.69	4.90	9.79	4.88	1.61	3.28	9.81	***	3.29	***	6.51	***
Senior Styles	14.86	2.25	12.61	6.50	1.91	4.59	8.36	***	0.34	NS	8.02	**
Rustic Outposts	16.37	2.68	13.69	4.48	1.49	2.99	11.89	***	1.19	NS	10.70	***
Midtown Singles	13.06	4.28	8.78	4.60	1.75	2.85	8.46	***	2.53	**	5.93	***
Hometown	13.41	3.50	9.91	5.41	1.77	3.65	8.00	***	1.73	NS	6.26	***
Next Wave	9.97	2.66	7.31	3.88	2.87	1.01	6.09	***	-0.21	NS	6.30	***
Scholars and Patriots	10.34	4.60	5.75	5.17	3.88	1.29	5.17	NS	0.72	NS	4.46	+
Full Sample	13.57	4.08	9.49	5.26	2.26	2.99	8.32	***	1.82	***	6.50	***

Note: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, NS = not significant. For the sake of space and readability, we do not provide detailed chi-square statistics and degrees of freedom, but this information is available upon request.

Another way to understand the effect of the $5 + priority mailing is to break down the differences in response between the treatment and control groups by mode and then compare how these differences vary across Esri Tapestry *LifeMode* groups. On average, sending the additional mailing significantly increased the web response rate by 1.82 p.p. and the mail (paper) response rate by 6.50 p.p., relative to the control group (see bottom of Table 1). The increases in web and paper responses were not consistent across Esri groups. Specifically, among *Uptown Individuals*, the web response rate was increased by a sizeable 4.44 p.p. ($p = 0.031$) but no effect was observed for the mail response rate (-1.18 p.p, $p = 0.553$), suggesting that the additional

paper questionnaire might not be necessary for this specific subgroup. For *Scholars and Patriots*, the web and mail response rates were increased by 0.71 (p = 1.00) and 4.45 p.p. (p = 0.06), respectively, relative to the control group, suggesting that the additional follow-up had a small effect on this subgroup, regardless of mode. Finally, for *Rustic Outposts*, the strong effect of this mailing was largely because it increased the mail response rate by 10.7 p.p. ($p < 0.001$) relative to the control group. For these types of rural areas, the additional mail screening questionnaire combined with the additional incentive proved quite effective.

As for the remaining *LifeMode* groups, the $5 + priority mailing always led to a significant increase in the mail response rate. But its effect on eliciting more web responses was less consistent. In half of these groups, the increase in the web response rate was small and non-significant. Thus, most of the screening questionnaires returned in response to the $5 + priority mailing were on paper. In general, the effectiveness of the $5 + priority mailing nonresponse follow-up did seem to vary across different types of geographic areas.

Hypothesis 2: Mail and Telephone Reminders Will Improve Survey Quality at the Main Stage. Turning next to the main stage of data collection, in the "full" condition, 12% of respondents participated by mailing back the shortened paper questionnaire. We found no significant differences between web and mail respondents in terms of gender, age, race and ethnicity, and income. We did find that mail consistently brought in individuals with lower education at a higher rate (59% of mail respondents had some college or less, compared to only 44% of web respondents; p = 0.044). Because the AFHS generally underrepresented persons with lower education (based on the benchmark NSFG, an estimated 65% of persons had some college or less

education), the mail follow-up contributed to improving representativeness in terms of education. These results were not sensitive to the inclusion of partial respondents.

Second, for the modular condition, 11%, 8% and 7% of respondents used the paper questionnaire in the three modules, respectively. The same education differences introduced by mail for the full condition emerged again. In module 1, module 2, and module 3, the proportions of some-college-or-less-educated respondents were 48%, 46%, and 45% among those who responded via web, respectively, compared to 69%, 76%, and 76% among those responded via paper questionnaire. Thus, there is a robust overall effect of the mail follow-up in terms of bringing in more individuals with lower education. These results were once again not sensitive to the inclusion of partial respondents.

Finally, we assessed whether the use of a telephone follow-up reminder at the final phase of the main data collection protocol served to increase response rates for different socio-demographic subgroups. The telephone reminders were generally quite effective at increasing response rates in both experimental conditions (+16.6 p.p. in module 1, +16.3 p.p. in module 2, +13.9 p.p. in module 3, and +13.7 p.p. in the full condition; all significant at $p < 0.001$ based on binomial tests in which the alternative hypothesis was no increase). In contrast, the mail (and email, if provided) follow-up reminders prior to the telephone reminders only tended to increase response rates by 3-7 p.p. across the conditions; the one exception was those invited to complete module 3, where the first follow-up reminder increased the main response rate by 11.7 p.p.

The telephone reminders were particularly effective for the non-Hispanic other female 20-49 group in module 1 (where the response rate increased by 15.2 p.p.) and the Black female 20-49 group in module 1 (+19.2 p.p.), module 2 (+29.8 p.p.), and module 3 (+17 p.p.). This suggests that a combination of modular design and telephone reminders may work well for older Black females. In sum, the use of telephone reminders at the final phase of the main data collection proved to be well worth the cost of this aspect of the sequential mixed-mode protocol. In general, we found robust support for our second hypothesis.

Hypothesis 3: Modular Survey Design Will Increase Main Survey Completion Rates.

We now evaluate whether the modular design approach increased response and completion rates overall. For the main instrument, the full approach resulted in a response rate of 62.9% (fully completed surveys), and 66.6% when including partial respondents. For the modular approach, only 42.9% of respondents completed all three modules with no partial responses ($p < 0.001$). Of those invited to complete module 1, 64.3% completed the module (not significantly higher than the full response rate overall), and this response rate was 65.4% when including partial respondents. Of the respondents who completed module 1, 82.6% completed module 2, and of those respondents, 79.2% went on to complete module 3. Given these results, the attrition in completing the entire survey that we found in the modular condition appeared to come from the additional requests to complete modules 2 and 3 (which is consistent with the prior literature in this area that was reviewed earlier). We therefore did not find support for our third hypothesis.

Next, we explored whether the modular approach tended to be more effective for particular types of individuals at the main stage. We found that the modular approach was particularly ineffective

for non-Hispanic other males between the ages of 20-49, where the completion rate in the full condition was 69.3%, and the modular completion rate was 42.1% (including partials, $p < 0.001$). The same was true for Black and Hispanic males between the ages of 20-49, where the differences were 53.1% vs. 26.7% ($p < 0.05$) and 59.3% vs. 36.7% ($p < 0.05$), respectively (including partials). The modular approach simply did not seem to work for adult males. The same pattern held up for other (73.7% vs. 49.4%, $p < 0.001$) and Hispanic (58.9% vs. 38.0%, $p = 0.019$) females between the ages of 20-49.

One interesting difference was for Black females between the ages of 20-49. For this subgroup, the module 1 response rate (78.0%) was significantly higher than the full response rate (55.6%, $p < 0.05$). The percentage of individuals in this subgroup who completed all three modules (56.0%) was still slightly higher than the full response rate among those who completed the screening interview, although this difference was not significant. This may be worth investigation in future adaptive designs. The module 1 difference did not hold up as significant when including partial respondents, but the difference was still notable. This was the only change in the pattern of results when including partial respondents.

Discussion

Summary of Findings

As we hypothesized, offering a short mail questionnaire to difficult non-respondents at later phases of both the screening and main stages of the self-administered AFHS data collection tended to recruit different types of respondents, specifically bringing in more respondents with lower education. At the screening stage, this approach also recruited more older respondents. We

also found support for our hypothesis that the use of a telephone follow-up reminder to complete the survey online or by mail with nonrespondents at the main stage of data collection (in this case at least six weeks out from the initial invitation to complete the main survey) would increase response rates at this important stage. This approach was especially effective, increasing response rates substantially in all conditions and subgroups relative to the mail and email nonresponse follow-up efforts.

Revisiting our first hypothesis, we also found that following up with "difficult" non-respondents late in the screening stage of this type of national web/mail data collection (e.g., 28 days after the initial invitation to complete the screener) and offering a small additional incentive ($5) along with an option to complete the screening questionnaire by mail was particularly effective at increasing response rates (as we hypothesized). The effectiveness of this approach does tend to vary across different types of geographic areas, and it appears to work especially well in rural areas like *Rustic Outposts*. For most types of areas (*Uptown Individuals* being an exception), the benefit of this approach arises from bringing in more screening responses by mail (rather than by web). As we noted earlier, attaching this type of commercial information to an address-based sampling frame is also relatively straightforward in different countries [56, 59]. Given these results, adaptive survey designs applying unique protocols to different subgroups defined by these commercial data warrant additional attention in the future.

With respect to the ability of modular survey design to increase perceived respondent convenience and reduce burden, we did not find support for our third hypothesis. Splitting a 60-minute web survey into three smaller modules (with completion times ranging between 5 and 20

minutes, depending on a respondent's characteristics) and offering a two-week break in-between the modules significantly reduced overall survey completion rates relative to asking the respondents to complete the survey online in one sitting and take breaks as needed. The reduction in the completion rate for the modular approach came primarily from lower response rates to the follow-up modules. Only for Black females aged 20-49 did this type of modular approach prove to be effective; reasons for this are unclear, and future research would benefit from more qualitative work to understand why this approach may be more effective for certain subgroups.

Overall, these findings are largely consistent with the relatively nascent literature in this area and provide more compelling evidence of the benefits of sequential mixed-mode design in these types of national web/mail surveys, where in-person interviewing is not an option for nonresponse follow-up and both mail and telephone follow-up efforts play important roles in increasing response rates and recruiting different types of respondents.

Implications for Practice

Regarding our main findings, we did see evidence of heterogeneity in the effectiveness of these approaches at both the screening stage (where we had linked information from the Esri Tapestry Segmentation about the type of area where a sampled person was living) and the main stage (where we had socio-demographic information available from the screening questionnaire). These findings for specific subgroups have important implications for future adaptive survey designs in these types of national web/mail data collections.

For example, when following-up with non-responding sampled households in wealthier urban areas with younger adults, the $5 + priority mailing used as part of the screening stage may not need to include a mail screener (see also [43]). As another example, using a modular design approach for older Black females at the main data collection stage may work well, but this finding needs replication in other settings. We found that linking the Esri Tapestry Segmentation data for our sampled areas to our sampling frame was a useful tool for evaluating areas where these approaches may have been more or less effective at the screening stage. These Esri segments are readily available, can easily be linked to address-based samples, and offer one way to cluster the distinctive features of neighborhoods based on data from 2010 Census, the American Community Survey, Esri's demographic updates and consumer surveys such as the Survey of the American Consumer ([55]; see also [56] for an example of this approach in another country). Future studies could use our results to tailor their screening approaches to particular types of areas defined by these data.

We did not find evidence of the modular data collection approach being effective at increasing response rates at the main data collection stage. However, we acknowledge that this approach was akin to asking respondents randomly assigned to this condition to complete three separate surveys, rather than a single survey (even though the modular surveys were one-third of the length, in an effort to decrease perceived burden). In short, this study suggested that asking respondents to complete several shorter web/mail surveys was not as effective as asking respondents to complete a single long web/mail survey and take breaks as needed; the number of survey requests, and not the length, seemed to be the limiting factor.

Overall, for future national surveys requiring household screening, we would recommend the following design strategies based on the results of this study:

1) the use of push-to-web approaches at the screening stage, where nonrespondents to the web invitation are mailed a paper questionnaire and a small additional pre-paid incentive;
2) the provision of shorter mail questionnaires for nonrespondents at the main stage;
3) the use of telephone reminders after the mail follow-up efforts at the main stage; and
4) allowing respondents to complete the entire web survey in a single sitting and take breaks as needed.

Slight modifications of these overall suggested approaches may be possible based on our findings for particular subgroups, and replications in other settings would be helpful for supporting these modifications. Studies with more financial resources may also consider the use of face-to-face follow-up if the telephone reminders do not prove effective; the AFHS did not have the resources to implement in-person interviewing at any stage of the study.

Future Research Directions

First, our $5 + priority mailing approach used to follow up with nonrespondents at the screening stage did not enable decomposition of what was most effective: the $5, or the priority mailing? In a second national replicate of the AFHS that was recently completed, we attempted to decompose this effect with a designed experiment, and analyses currently in progress suggest that the additional $5 and the priority mailer have *additive*, rather than *multiplicative*, effects on the response rate [60]. Replications of this finding are needed in other contexts.

Second, we still feel that further evaluation of the modular design approach in web surveys is worthwhile. The 20-minute survey request in this study may have negated some of the possible benefits of the modular approach, and shorter questionnaires combined with micro-incentives may still prove valuable for increasing perceived convenience [38]. Future experimental work with these kinds of approaches would be valuable, but this study suggests that the modular approach may not work well with longer, complicated surveys.

Third, our screening response rate was much lower than expected given the prior literature in this area and similar screening efforts from other national surveys prior to the COVID-19 pandemic. All the data collection for this replicate of the AFHS took place during the pandemic, and much of the data collection took place during the final year of the Trump administration. This context needs to be considered when evaluating the response rates, as a sensitive political climate combined with distrust in the government and/or science may have affected the response rates for this NIH-sponsored survey [61]. We will continue to evaluate response rates in our second national sample replicate of the AFHS, but ongoing reporting of response rates to these types of web/mail screening efforts in other contexts would be welcome from other survey organizations to see if these response rates were part of a larger societal pattern.

Finally, this paper did not address the effects of these alternative approaches on substantive responses to the main survey, or on other paradata related to the approaches (e.g., response times, item missing data, etc.). We are actively working on these evaluations, which generally provide support for the ability of the AFHS approach to replicate findings based on the NSFG despite the lower response rate [62], but future work needs to consider these substantive and paradata-

related outcomes as well before the field can truly decide on optimal approaches for these types of national web/mail surveys.

Conclusion

The design recommendations supported by this study will make it possible for researchers with limited resources to conduct robust national surveys, including the launch of cross-sectional surveys on new topics that will directly benefit the well-being of a larger population. These recommendations would also support the timely and cost-efficient implementation of high-quality surveys in smaller geographic areas, such as specific U.S. states or cities. The new national survey data collection design that we recommend is a cumulative combination of important recent advances in survey methodology. The analyses presented here validate this recommended methodological approach, but they also document the limitations of this new alternative to large-scale face-to-face surveys. Continued evaluations of this recommended approach in different settings internationally would help to further validate the potential of the approach to improve the quality of large-scale surveys. In addition, more widespread use of evidence-based adaptive survey designs will help to tailor these recommended approaches to specific subgroups, ideally increasing response rates and reducing the nonresponse bias in survey estimates of interest.

Acknowledgements

The authors wish to acknowledge the contributions of Mick Couper, Colette Keyser, Andrew Hupp, and the rest of the AFHS team in the Survey Research Operations unit of the Survey Research Center.

References

1. Tourangeau R, Edwards B, Johnson TP, Wolter KM, Bates N. Hard-to-survey populations: Cambridge University Press; 2014.

2. Tourangeau R, Conrad FG, Couper MP. The science of web surveys: Oxford University Press; 2013.

3. Couper MP. Designing effective Web surveys: Cambridge University Press; 2008.

4. Couper M, Antoun C, Mavletova A. Mobile Web Surveys. In: Biemer P, Leeuw Ed, Eckman S, Edwards B, Kreuter F, Lyberg L, et al., editors. Total Survey Error in Practice. New York: Wiley; 2017. p. 133-54.

5. Lugtig P, Toepoel V. The use of PCs, smartphones, and tablets in a probability-based panel survey: Effects on survey measurement error. Social Science Computer Review. 2016;34(1):78-94.

6. Mavletova A, Couper MP. Sensitive topics in PC web and mobile web surveys: is there a difference? Survey Research Methods. 2013;7(3):191-205.

7. Couper MP. Is the sky falling? New technology, changing media, and the future of surveys. Survey Research Methods. 2013;7(3):145-56.

8. Couper MP, Gremel G, Axinn W, Guyer H, Wagner J, West BT. New options for national population surveys: The implications of internet and smartphone coverage. Social Science Research. 2018;73:221-35.

9. Manfreda KL, Bosnjak M, Berzelak J, Haas I, Vehovar V. Web surveys versus other survey modes: A meta-analysis comparing response rates. International Journal of Market Research. 2008;50(1):79-104.

10. Shih T-H, Fan X. Comparing response rates from web and mail surveys: A meta-analysis. Field Methods. 2008;20(3):249-71.

11. Groves RM, Singer E, Corning A. Leverage-saliency theory of survey participation: description and an illustration. Public Opinion Quarterly. 2000;64(3):299-308.

12. Dillman DA, Smyth JD, Christian LM. Internet, phone, mail, and mixed-mode surveys: The tailored design method: John Wiley & Sons; 2014.

13. Olson K, Smyth JD, Horwitz R, Keeter S, Lesser V, Marken S, et al. Transitions from Telephone Surveys to Self-Administered and Mixed-Mode Surveys: AAPOR Task Force Report. Journal of Survey Statistics and Methodology. 2020;9(3):381-411.

14. Axinn WG, Gatny HH, Wagner J. Maximizing data quality using mode switching in mixed-device survey design: nonresponse bias and models of demographic behavior. Methods, Data, Analyses 2015;9(2):163.

15. Millar MM, Dillman DA. Improving response to web and mixed-mode surveys. Public Opinion Quarterly. 2011;75(2):249-69.

16. Singer E. The use of incentives to reduce nonresponse in household surveys. Survey Nonresponse. 2002;51(1):163-77.

17. DeBell M, Amsbary M, Meldener V, Brock S, Maisel N. Methodology report for the ANES 2016 time series study. Palo Alto, CA, and Ann Arbor, MI: Stanford University and the University of Michigan; 2018.

18. Debell M. Nonresponse Bias in a Nationwide Dual-Mode Survey. International Total Survey Error Workshop; Durham, NC 2018.

19. Luiten A, Hox J, De Leeuw E. Survey nonresponse trends and fieldwork effort in the 21st century: results of an international study across countries and surveys. Journal of Official Statistics. 2020;36(3):469-87.

20. Williams D, Brick JM. Trends in US face-to-face household survey nonresponse and level of effort. Journal of Survey Statistics and Methodology. 2018;6:186-211.

21. Brick JM, Williams D, Montaquila JM. Address-based sampling for subpopulation surveys. Public Opinion Quarterly. 2011;75(3):409-28.

22. Montaquila JM, Brick JM, Williams D, Kim K, Han D. A study of two-phase mail survey data collection methods. Journal of Survey Statistics and Methodology. 2013;1(1):66-87.

23. Brick JM, Montaquila JM, Han D, Williams D. Improving response rates for Spanish speakers in two-phase mail surveys. Public Opinion Quarterly. 2012;76(4):721-32.

24. Han D, Montaquila JM, Brick JM. An evaluation of incentive experiments in a two-phase address-based sample mail survey. Survey Research Methods. 2013;7(3):207-18.

25. Biemer PP, Murphy J, Zimmer S, Berry C, Deng G, Lewis K. Using bonus monetary incentives to encourage web response in mixed-mode household surveys. Journal of Survey Statistics and Methodology. 2018;6(2):240-61.

26. Biemer P, Murphy J, Zimmer S, Berry C, Deng G, Lewis K, editors. A test of Web/PAPI protocols and incentives for the residential energy consumption survey. Annual Meeting of the American Association for Public Opinion Research, Austin, TX; 2016.

27. Zimmer S, Biemer P, Kott P, Berry C, editors. Testing a Model-Directed, Mixed Mode Protocol in the RECS Pilot Study. Proceedings of the 2015 Federal Committee on Statistical Methodology Research Conference; 2015.

28. Brick JM, Andrews W, Brick PD, King H, Mathiowetz NA, Stokes L. Methods for improving response rates in two-phase mail surveys. Survey Practice. 2012;5(3):1-7.

29. Mittereder F, West BT. A dynamic survival modeling approach to the prediction of web survey breakoff. Journal of Survey Statistics and Methodology. 2022;10(4):945-78.

30. Peytchev A. Survey Breakoff. Public Opinion Quarterly. 2009;73(1):74-97.

31. Johnson A, Kelly F, Stevens S. Modular Survey Design: Bite-Sized Chunks Proposal. CASRO Online Research Conference; San Fransisco, CA, 2013.

32. Andreadis I, Kartsounidou E. The impact of splitting a long online questionnaire on data quality. Survey Research Methods. 2020;14:31-42.

33. Peytchev A, Peytcheva E, Conzelmann JG, Wilson A, Wine J. Modular survey design: Experimental manipulation of survey length and monetary incentive structure. Journal of Survey Statistics and Methodology. 2020;8(2):370-84.

34. Toepoel V, Lugtig P. Modularization in an era of mobile web: investigating the effects of cutting a survey into smaller pieces on data quality. Social Science Computer Review. 2022;40(1):150-64.

35. West BT, Ghimire D, Axinn WG. Evaluating a modular design approach to collecting survey data using text messages. Survey Research Methods. 2015;9(2):111-23.

36. Johnson A, Kelly F, Stevens S. Modular Survey Design for Mobile Devices. CASRO Online Research Conference; Las Vegas, 2012.

37. Smith R, Kotzev I, Miller C, Kachhi D. Modularizing surveys to meet today's respondent challenges. 2012 CASRO Online Conference, 2012.

38. Bell AR, Ward PS, Killilea ME, Tamal MEH. Real-time social data collection in rural Bangladesh via a 'microtasks for micropayments' platform on android smartphones. PLoS ONE. 2016;11(11):e0165924.

39. Gatny HH, Couper MP, Axinn WG, Barber JS. Using debit cards for incentive payments: Experiences of a weekly survey study. Survey Practice. 2009;2(7):2945.

40. Sugie NF. Utilizing smartphones to study disadvantaged and hard-to-reach groups. Sociological Methods & Research. 2018;47(3):458-91.

41. Ballivian A, Azevedo JP, Durbin W, Rios J, Godoy J, Borisova C. Using mobile phones for high-frequency data collection. Mobile Research Methods. 2015;21.

42. Biemer P, Harris K, Burke B, Considine K, Halpern C, Suchindran C, editors. The transition of an in-person panel survey to a web-mail mixed mode design. Methodology of Longitudinal Surveys II Conference; 2018; Essex, UK.

43. Biemer PP, Harris KM, Burke BJ, Liao D, Halpern CT. Transitioning a panel survey from in- person to predominantly web data collection: Results and lessons learned. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2021.

44. Wolf C, Christmann P, Gummer T, Schnaudt C, Verhoeven S. Conducting general social surveys as self-administered mixed-mode surveys. Public Opinion Quarterly. 2021;85(2):623-48.

45. Peytchev A, Peytcheva E. Reduction of measurement error due to survey length: Evaluation of the split questionnaire design approach. Survey Research Methods. 2017;11(4):361-8.

46. Ioannidis E, Merkouris T, Zhang L-C, Karlberg M, Petrakos M, Reis F, et al. On a Modular Approach to the Design of Integrated Social Surveys. Journal of Official Statistics (JOS). 2016;32(2).

47. Gonzalez JM, Eltinge JL, editors. Multiple matrix sampling: A review. Proceedings of the Section on Survey Research Methods, American Statistical Association; 2007: American Statistical Association Alexandria, VA.

48. Raghunathan TE, Grizzle JE. A split questionnaire survey design. Journal of the American Statistical Association. 1995;90(429):54-63.

49. Luijkx R, Jónsdóttir GA, Gummer T, Ernst Stähli M, Frederiksen M, Ketola K, et al. The European Values Study 2017: On the way to the future using mixed-modes. European Sociological Review. 2021;37(2):330-46.

50. Beckett MK, Elliott MN, Gaillot S, Haas A, Dembosky JW, Giordano LA, et al. Establishing limits for supplemental items on a standardized national survey. Public Opinion Quarterly. 2016;80(4):964-76.

51. Burkhart Q, Orr N, Brown JA, Hays RD, Cleary PD, Beckett MK, et al. Associations of mail survey length and layout with response rates. Medical Care Research and Review. 2021;78(4):441-8.

52. Galesic M, Bosnjak M. Effects of questionnaire length on participation and indicators of response quality in a web survey. Public Opinion Quarterly. 2009;73(2):349-60.

53. Robb KA, Gatting L, Wardle J. What impact do questionnaire length and monetary incentives have on mailed health psychology survey response? British Journal of Health Psychology. 2017;22(4):671-85.

54. Wagner J, West BT, Couper MP, Zhang S, Gatward R, Nishimura R, et al. An Experimental Evaluation of Two Approaches for Improving Response to Household Screening Efforts in National Mail/Web Surveys. Journal of Survey Statistics and Methodology. 2022.

55. ESRI. Tapestry Segmentation: Redlands, CA: ESRI; 2021 [Available from: https://doc.arcgis.com/en/esri-demographics/latest/regional-data/tapestry-segmentation.htm#ESRI_SECTION1_87F5D845F8E04723AE1F4F502%20FF3B636].

56. West BT, Kreuter F, Trappmann M. Is the collection of interviewer observations worthwhile in an economic panel survey? New evidence from the German Labor Market and Social Security (PASS) Study. Journal of Survey Statistics and Methodology. 2014;2(2):159-81.

57. Mulry MH, Bates N, Virgile M. Viewing participation in censuses and surveys through the lens of lifestyle segments. Journal of Survey Statistics and Methodology. 2021;9(4):764-88.

58. Bucks B, Couper MP, Fulford SL. A mixed-mode and incentive experiment using administrative data. Journal of Survey Statistics and Methodology. 2020;8(2):352-69.

59. West BT, Wagner J, Hubbard F, Gu H. 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. 2015;3(2):240-64.

60. Zhang S, West BT, Wagner J, Couper MP, Gatward R, Axinn WG. Visible cash, a second incentive, and priority mail? An experimental evaluation of mailing strategies for a screening questionnaire in a national push-to-web/mail survey. Journal of Survey Statistics and Methodology (Revise and Resubmit). June 2022.

61. Sastry N, McGonagle K, Fomby P. Effects of the COVID-19 crisis on survey fieldwork: experience and lessons from two major supplements to the US Panel Study of income dynamics. Survey Research Methods. 2020;14(2):241.

62. West BT, Couper MP, Axinn WG, Wagner J, Gatward R, Saw H, Zhang S. Evaluating a Web/Mail Alternative to a National Face-to-Face Survey: Initial Results from the American Family Health Study. Paper presented at the 2022 Conference of the American Association for Public Opinion Research (AAPOR), May 12, 2022, Chicago, IL.

Supporting Information

S1 Text. Full Description of ESRI Tapestry Segmentation.

S2 Text. AFHS Response Rate Calculations.

Figure 1

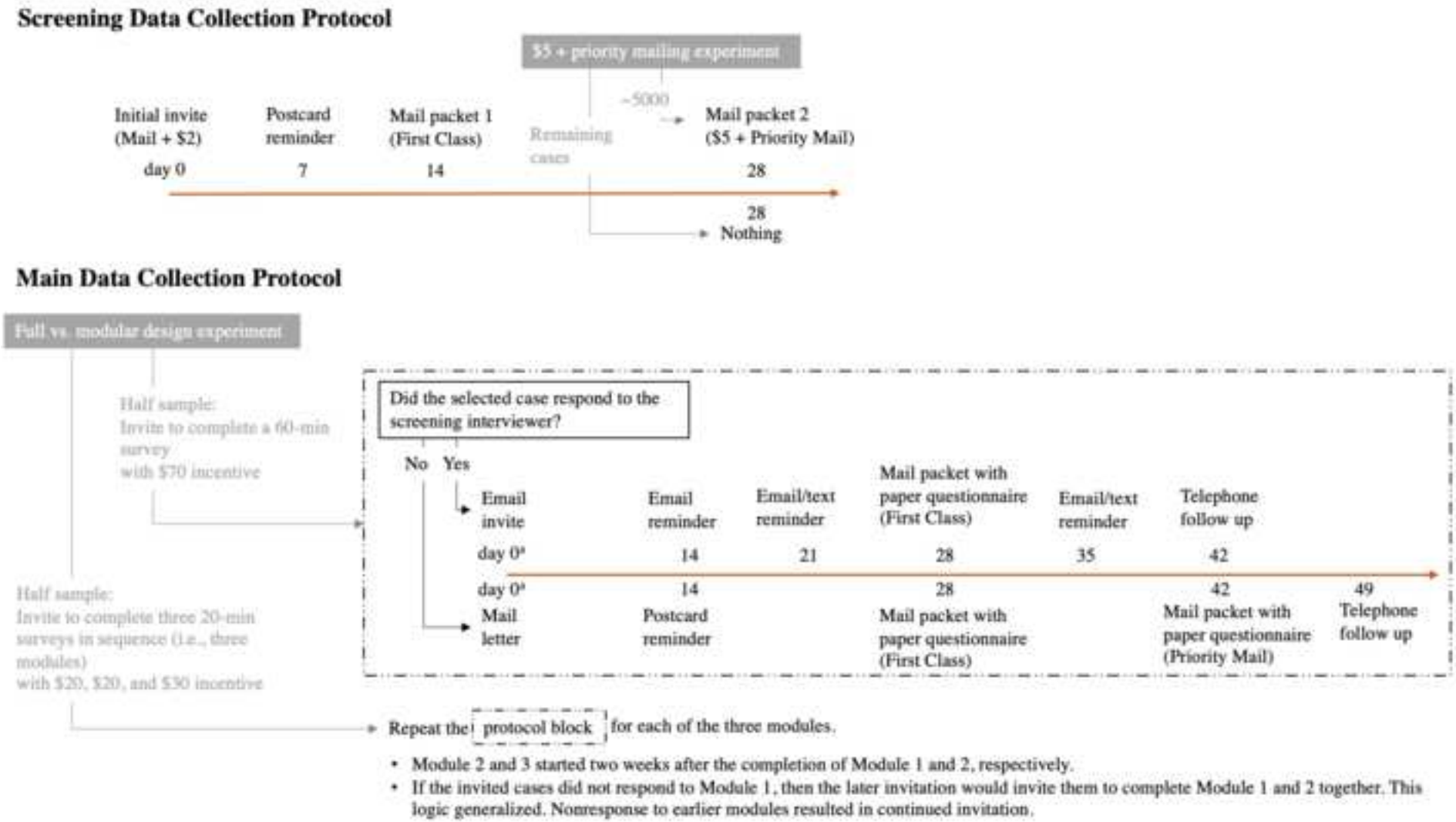

Figure 2

Figure 3