

AFHS Methodology

March 30, 2020

AFHS Sample Selection. The AFHS will screen around 42,120 randomly sampled U.S. households either online (individuals in sampled households will initially be encouraged to complete a web version of the screening questionnaire) or by mail (via follow-up of initially sampled households that do not respond online) to identify eligible household members between the ages of 18 and 49 (the current NSFG target population, excluding children who are between the ages of 15 and 17). We propose to divide this screening work across two successive 9-month “replicates,” where each replicate is a fully representative sample of U.S. households. One randomly selected eligible respondent within each household completing the screening questionnaire will subsequently be invited to complete the three aforementioned modules of content, either online or by returning paper questionnaires. This plan (see *Data Collection*) is designed to obtain at least one returned module from 4,000 respondents.

We worked with the GENESYS system of Marketing Systems Group (MSG) to select address-based samples from each of the 50 U.S. states and Washington D.C. Overall, across both replicates, the AFHS plans to select approximately 42,120 addresses using probability sampling, of which we expect approximately 18,820 returned screening interviews (or about 9,410 per replicate). Of these returns, we expect approximately 10,754 addresses to contain an eligible person (or 5,377 eligible persons per replicate; see *Data Collection*). For a given replicate, we will first allocate the sample of 21,060 addresses proportionately across the U.S. states and Washington D.C. based on the most recent American Community Survey five-year data. This will ensure sample representation from each state and major region of the U.S.

Next, we requested that MSG further stratify the allocated samples for each state based on auxiliary information that MSG makes available for households in these types of address-based samples. Our previous research in this area has indicated that auxiliary information on the ages of household members can be quite useful in improving the efficiency of screening operations based on age. We therefore created three strata based on this auxiliary information: 1) likely eligible, 2) likely ineligible, and 3) no information available. Because there are eligible persons in each of these strata, we will sample from each of them. We oversampled from the stratum of likely eligible cases and undersampled from the stratum of likely ineligible cases. We also defined an optimal allocation of sample to the three strata. This optimal allocation raises the expected eligibility rate from about 0.47 (under an allocation of sample to the strata that is proportional to the stratum sizes) to about 0.57. We also oversampled housing units from Census Block Groups with more than 10% African-American or Hispanics. This results in oversamples of African-Americans and Hispanics such that about 20% of completed surveys are from members from each of these groups. This process will result in stratified random samples of addresses from each of the 50 states, with an oversample of the sampled addresses in each state selected from households that are indicated as having someone between the ages of 18 and 49 present and from neighborhoods with more than 10% African-Americans or Hispanics.

An invitation to complete a screening questionnaire online will then be sent to the sampled addresses in each state, and we will follow-up with initial non-respondents to the screening invitation using a sequential mixed-mode approach (see *Data Collection*). A single respondent will be randomly selected from households completing the screening questionnaire that contain more than one eligible person; this will occur immediately for screening questionnaires completed online, and after scanning and processing for screening questionnaires completed via mail. The within-household selection rates will be varied such that women and teens will be oversampled relative to their proportions in the population. Collectively, these procedures emulate the current sampling procedures of the NSFG. In general, we aim to mimic the sampling procedures used by the NSFG to the greatest extent possible given budget constraints so that we can evaluate the survey results produced under an alternative and more cost-efficient data collection protocol.

Once we have randomly selected eligible persons for one of the two replicates in the main study, they will be randomly assigned to either the modular condition (50%) or the single-sitting condition (50%) for the main data collection. If an individual completes the screening questionnaire for a household online and is also the individual from that household randomly selected for the main study, that individual will immediately be randomized to one of these two conditions and invited to continue completing either the full study questionnaire or the first module of the study online (depending on the condition to which they have been assigned). If that individual chooses not to continue with the main questionnaire at this time, we will provide a “thank you” message and indicate that we will follow up with them via mail at a subsequent date to proceed with the main study.

For all respondents randomly assigned to the modular condition, we will invoke our tracking systems to follow each case through the process of each data collection module. The Michigan team implementing this project is at the forefront of developing and evaluating new methods of interviewing. This includes design of state-of-the-art tracking systems for mixed-mode surveys. This tracking will ensure that we have the most up-to-date contact information for follow-up effort, including addresses and telephone numbers, maintaining security of confidential respondent data throughout the process. We propose to use the University of Michigan Survey Research Center’s **new mixed-mode sample management system** (or SMS) for monitoring the web survey completion rates and tracking follow-up steps (mailing, email, text reminders, and telephone calls; see **Data Collection**) on all questionnaire components. This sample management system will create customized dashboards for tracking production goals and cost information. The data from our sample management system will be continuously analyzed, providing the means to optimize reminders and related protocol steps on a daily basis and in the second replicate.

Interviewer Training. Although AFHS data collection will occur primarily online with backup by mail, we will use modest effort from professional interviewers to conduct reminder calls (given phone numbers collected in screening; see below), send text messages, and answer respondent questions. All interviewers are expected to work from our centralized facility in Ann Arbor. On-staff interviewers will receive half a day of training, which will involve reviewing the data collection protocols for this project, making non-response / reminder calls, and using the tailored sample management system features. The interviewer training will be held in Ann Arbor.

AFHS Data Collection Procedures. Data collection will begin with our **screeener data collection** before main data collection begins. The screener protocols are designed to fit within **four weeks**, and include the following:

- An initial invitation to complete the screener questionnaire online, including a cover letter describing the study and the initial invitation (with a customized web link to the online screening questionnaire), and a \$2 token of appreciation attached to the top of the letter.
- **A Spanish-Speaking Household Experiment.** Prior to the onset of screener data collection, we will build statistical models for an indicator of whether a sampled NSFG household has a Spanish-speaking individual present (using historical NSFG data). An estimated 4% of NSFG interviews are completed in Spanish, and we will compute predicted probabilities of households having a Spanish-speaking individual present for all sampled AFHS households, using MSG data and sampling frame information as predictors. All households with predicted probabilities that lie above a pre-determined threshold in terms of their predicted probability (based on an ROC curve analysis, designed to maximize the sensitivity and specificity of decisions based on this threshold) will be randomized to one of two conditions: a two-sided cover letter (one side English, one side Spanish), or an English cover letter only. We hypothesize that the two-sided letter will increase rates of response to the screening questionnaire for this specific subgroup of sampled households. Of the cases that do not lie above the pre-determined

threshold, 90% will be randomized to receive the English cover letter only, and 10% will be randomized to receive the two-sided cover letter. This will enable an experimental assessment of the benefits and/or drawbacks of the two-sided letter for both subgroups of sampled addresses. For respondents completing the screening questionnaire, the initial (or follow-up) paper mailings requesting their participation in the main portion of the data collection (Module 1 or the full instrument) will be provided in both English and Spanish. For those in the modular condition, we will later tailor these paper mailings to their language of choice based on the first completed module.

- The online screening questionnaire will request respondents to roster the household, which will include providing relationship status, age, gender, and race / ethnicity.
- The screening questionnaire will also request an email address and a telephone number for the household informant, permission to contact the respondent by text message (if applicable), and information about internet access.
- Exactly three days after the initial screener invitation, a sealed reminder postcard will be mailed to all selected households.
- Approximately 1 week following the thank you/reminder postcard, a screener questionnaire packet with a different cover letter (targeted to initial non-respondents), a paper English screener questionnaire, a paper Spanish screener questionnaire, and a postage-paid return envelope will be sent to all active non-respondents using USPS first class mail.
- Approximately 14 days following the second screener invitation, a third and final screener invitation packet (with a different [final] cover letter, a \$5 pre-paid token of appreciation, and the same materials otherwise) will be mailed to all non-respondents using USPS Priority Mail.

The proposed four-week screener period encompasses all mail out-mail back lags associated with this plan. Overall, based on published response rates for similar mailing protocols, we estimate that this plan will produce a screener phase response rate of approximately 45%. The National Household Education Survey has a similar design to the one proposed here. They used a mail screener followed by a web/mail administration of the survey. In a large 2011 field test, they obtained a 69% response rate for the screening survey. They tested \$2 versus \$5 incentives and found that the \$2 incentive for the screener obtained a 67% response rate.

Next, we will apply the **main data collection** protocols. Once the screener questionnaires are submitted / returned and processed / scanned, we will apply an automated respondent selection protocol to the household screener information. The selection protocol, which will occur instantly for all online submissions, will be 100% consistent with the current NSFG within-household selection protocol, resulting in over-samples of women and 18-19 year-olds.

Respondents selected for the main study will be randomly assigned to either the modular condition (50%) or the single-sitting condition (50%). As was mentioned above, any individuals completing the screener questionnaire online who are also randomly selected from their households will be immediately assigned to one of the two conditions and invited to either complete the first module online, or the full survey online (depending on their assignment). All other individuals will receive a thank-you message, and will be mailed an invitation letter to either complete the first module online (for a \$20 post-paid incentive) or to complete the full survey online (for a \$70 post-paid incentive). These respondents will receive a customized link in this invitation letter, making it easy to begin the survey. Individuals completing either the first module or the full survey will be mailed a check containing their incentive payment.

Nonresponse Follow-up Protocol. After two weeks, if no response to either the first module or the full survey (depending on assignment) is received, respondents will receive an initial follow-up contact, either by email (if they completed the screening questionnaire and provided an email address) or by regular mail (if the individual was not the one who completed the screening questionnaire and/or they did not provide an email address). For those providing an email address or cell phone number in the completed screening questionnaire, they will receive an additional email reminder (and / or text message reminder, if they indicated that study-related text messages to the cell phone number provided would be acceptable) one week later. For those receiving mail, they will receive a first class USPS mailing including a cover letter reminder and a paper version of either the first module or the full questionnaire (with a postage-paid return envelope) two weeks later. The group providing an email address that has not yet responded will receive the same mail package at the same time. The “mail” group (with no email address) will receive a second mail package two weeks after that; the “email” group will receive another email / text reminder encouraging completion one week later. Finally, subsamples of the “email” group predicted to have higher response propensity after six weeks will receive follow-up telephone call reminders, given the presence of a telephone number in the screener questionnaire or the MSG data. A similar subsample of the “mail” group will receive similar telephone calls starting seven weeks after the initial invitation letter was mailed. Up to seven calls will be attempted for each non-responding household in these subsamples, and in the modular condition, the same nonresponse follow-up protocol will be applied for each module (see details below).

All respondents completing the Module 1 questionnaire will be mailed a \$20 token of appreciation on a rolling, weekly basis. This payment process will be repeated for each of the three modules, with payments of \$20 and \$30 being sent if modules two and three are submitted, respectively. For Module 1 respondents, an invitation letter to complete Module 2 will be mailed **two weeks** after completion of the first module (on a rolling basis). The same two-week break will apply to the time between Module 2 and Module 3. If an individual assigned to the modular condition does not complete Module 1 after the full nonresponse protocol described above has been applied, they will receive an invitation to complete both Modules 1 and 2 (for \$40) after two weeks have elapsed. The same idea will hold for an individual who completes Module 1, but not Module 2; they will be offered both Module 2 and Module 3 two weeks after the end of the nonresponse protocol for Module 2 (for \$50). If an individual does not respond after the entire nonresponse protocol for both Module 1 and Modules 1 and 2, they will not be contacted further.

Overall, the two-replicate design will allow AFHS to respond to measurement and nonresponse issues identified in the first replicate. For example, if response rates to follow-up modules are lower than expected, we can experiment with offering a higher incentive or “bonus” for the completion of all modules in the second replicate. Another option would be to identify variables that are highly predictive of other survey content and ask these questions in the first module. These variables could then be used to impute item-missing values in later modules. If we identify item nonresponse as an issue on particular items or subsets of items, we could also experimentally implement strategies to address this problem in the second replicate. For example, we could add prompts to these specific items for a random half of the sample in the second replicate that give further assurances of confidentiality and stress the importance of supplying complete and accurate data.

We acknowledge that the mail mode of data collection does not compete with CATI or CAPI follow-up in terms of automated skip pattern implementation or auto-fill of past responses. However, our primary aim is to evaluate the ability of these more cost-effective alternatives to replicate current NSFG estimates, and our primary efforts will be to push as many respondents as possible to respond via web. The paper questionnaires will have limited skip logic and focus on measuring key NSFG items from individuals without web access, permitting the evaluation of possible selection bias in estimates arising from the web-only and mail-to-web protocols described above.

Post-Collection Processing / Public Availability of AFHS Data. Comparisons of estimates based on the survey data collected here with estimates based on public-use data files from the same time period for the NSFG will require that the same post-collection processing steps used by the NSFG be applied here. Specifically, this will entail the computation of respondent weights describing the probability sampling procedure. Respondents will have different probabilities of selection. First, households will have different selection rates based on the likelihood of eligibility determined from the auxiliary data as described earlier. Second, households in high-density African-American and Hispanic neighborhoods will have higher rates of selection. Third, the within-household selection rates will be differentiated to produce oversamples of teens and women. The overall probabilities of selection will be determined for each respondent using the same procedures presently used in the NSFG, and these probabilities will be inverted to form base sampling weights. Nonresponse adjustments will be developed using response propensity stratification. The logistic regression models used to estimate response propensity will include predictors similar to those used on the NSFG, including neighborhood characteristics available from the Decennial Census and American Community Survey, the commercial data described earlier, and information from household screenings, and these models will be estimated using the base sampling weights. The adjusted base weights will be calibrated to be in alignment with American Community Survey 5-year estimates of population control totals for post-strata defined by combinations of age, gender, and race-ethnicity. Finally, in order to allow users to combine data from this study with the public use files released by the NSFG (which has the significant benefit of increasing subgroup sample sizes), we will provide users with new respondent weights for a *combined* data set (our data, with common variable names, and the most recent NSFG data) that have been post-stratified to the same population control totals.

In addition to the development of final respondent weights for analysis, we will also retain codes describing the strata from which each respondent was sampled. These stratum codes will be used for variance estimation purposes in order to increase the efficiency of the estimates. A notable difference between this proposed design and the current face-to-face NSFG design is the absence of multi-stage cluster sampling, which generally reduces the efficiency of survey estimates. While cluster sampling would generally reduce the travel costs of a major national face-to-face data collection, we note that the increases in standard errors of survey estimates introduced by this cost-saving measure would not be needed under the proposed design (where no travel is needed). We therefore expect standard errors of weighted estimates based on the proposed design to generally be similar to or smaller than standard errors that will be computed using the public-use NSFG data (despite each study producing the same number of interviews), as design effects due to the multi-stage cluster sampling will be avoided (which is another key strength of the proposed methodology). We plan to employ Taylor Series Linearization for variance estimation in all analyses.

Finally, the standard set of edits applied to data collected using the present NSFG design will be programmed into the data collection systems used to process the web survey responses and the mail survey responses. Consistent with the current NSFG design, we will also impute item-missing values prior to analysis based on the final processed data sets. To enable integration of the data collected here with data from public-use NSFG files collected from a similar period, we will prepare code for combining a public-use NSFG data file from the same time period with our data set (to be made available via the ICPSR data archive), including the calibration of weights (as described above) and the calculation of corresponding stratum and cluster codes for our data (where each respondent is their own cluster). We will then provide data users with analysis syntax implementing state-of-the-art techniques for combining data from different surveys, where there may be mode effects (face-to-face vs. web/mail in our case) and one survey (e.g., NSFG) can be considered a “gold standard”.