

The National Health and Nutrition Examination Survey (NHANES) Incentive Pilot Details

Overview

A recent decline in response rates on several national face-to-face surveys has been documented¹. Survey response rate is a valuable data quality measure and the most widely used indicator of survey quality. A high response rate increases the likelihood that the survey accurately represents the target population. However, a lower response rate is not always associated with higher levels of nonresponse bias, and the levels of nonresponse bias can differ for different estimates in the same survey. Nonresponse bias can be substantial when two conditions hold: (1) the response rate is relatively low, and (2) the difference between the characteristics of respondents and nonrespondents is relatively large.

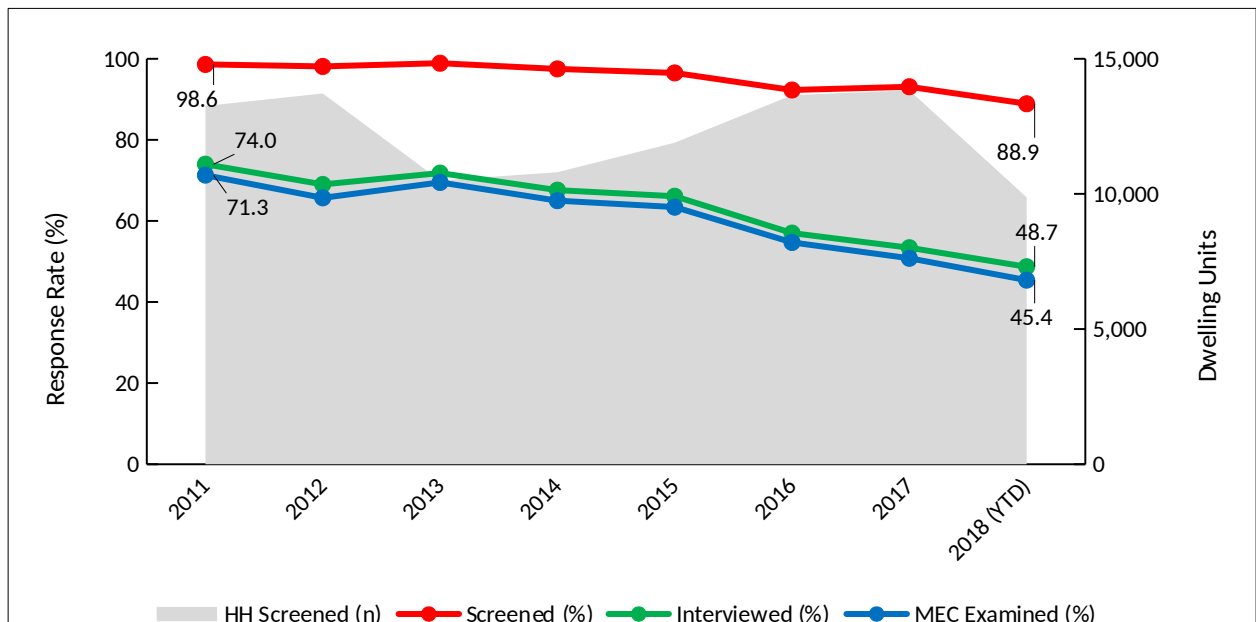
Since 2011, the National Health and Nutrition Examination Surveys (NHANES) has observed continuous decreases in its overall response rate¹. A review of each of the survey components (i.e. the screener, the in-home interview, and the MEC exam) has shown this decrease is largely occurring at the sample person (SP) in-home interview (see **Figure 1**). To better understand non-response bias on NHANES, and to characterize differences in the health characteristics of respondents and non-respondents, a non-response bias module will be added to the eligibility screening stage of the survey in 2019. The non-response bias module is composed of five health questions and can be found in **Appendix A**.

Currently, NHANES only offers respondents a monetary token of appreciation for completing the Mobile Examination Center (MEC) physical examination. This is done due to the high amount of burden required of respondents in terms of travel, time, and potential discomfort resulting from the physical examinations. Given the continued declines in response rates to all survey stages, and the growing challenges associated with collecting high-quality and unbiased nationally representative data, NHANES would like to test the use of incentives for the newly

added non-response bias module, which will be added to the screening stage in 2019, and to the SP in-home interview. The incentive experiment will be used to assess:

- 1) The impact of the incentives on data quality by reducing non-response bias in estimates;
- 2) The impact of the incentives on survey response rates, overall and by survey stage (i.e., the screener, the in-home interview, and the MEC exam); and
- 3) The impact of the incentives on the level of effort (i.e., contact attempts) required to obtain cooperation and the cost effectiveness of the incentives.

Figure 1. Response Rates by Survey Component and the Number of Screened Dwelling Units



Screener and Non-Response Bias Module

Screener Response Rates

The NHANES screener response rate for the most recent completed year (2017) is 93.1%. This is lower than previous years (3 percentage points lower than 2015 and 4 percentage points lower than 2014) and 2018 continues to decrease with 88.9% for the first ten stands of 2018. This is a cumulative decrease of nearly 9 percentage points since 2013 (an average drop of almost 2 percentage points per year). Literature reviews have shown that it is requiring more effort to obtain response (as measured in terms of number of contact attempts) even while response rates decline ^{1,2}. For 2017 and the first ten stands of 2018, the average number of contact attempts to

complete an NHANES screener with a household member was 4.1 (this excludes the two advance mail contacts informing households of a visit).

The recent and rapid pace of decrease in response combined with the high level of effort results in separate negative influences on data quality. Decreased response affects representation of different population subgroups. With decreased screener response, the number of households where the number of SPs is unknown increases. It is not possible to determine the composition of these households since they are nonrespondents, but the subgroup participation in the SP interview (conditional on screener completion) can be used as an indicator. Since 2015, Black (all race categories mentioned in this document are Non-Hispanic) and Hispanic SP conditional participation has decreased at a faster pace than white or Asian conditional participation indicating potential bias in group representation:

- Non-Hispanic white: 7 percentage point decrease (65.1% to 58.3%)
- Non-Hispanic black: 10 percentage point decrease (75.9% to 66.2%)
- Hispanic: 17 percentage point decrease (73.8% to 56.7%)
- Non-Hispanic Asian: 3 percentage point decrease (47.1% to 43.8%)

Screener response rates in the Primary Sampling Units (PSUs) completed in 2017 and the first ten stands of 2018 have ranged from 73.3% in one PSU to 98.5% in another PSU. Continued increases in differences between PSUs, and more frequent large differences between PSUs, will create more variance in the final weighted results, leading to a decrease in the precision of any published estimates.

NHANES spends 10 weeks at each stand (PSU location). During this time, interviewers must screen sampled dwelling units and conduct interviews for sampled household members, then the SP must complete the MEC exam after traveling to the MEC. Screening to identify SPs is conducted late into the stand field period. This leaves very little time to contact SPs who were not the screener respondent, conduct the SP interview and allow time for a MEC appointment convenient to the SP. For hard-to-reach minority subgroups, e.g., Hispanics and Asians, this results in under-representation. Efforts that can reduce this (e.g., reducing the number of

contacts) can improve representation and indirectly improve SP interview participation – by allowing more time to make contact and gain cooperation.

NHANES will also experience a further decline in screener response and an increase in effort expended (increase in contacts) due to a recent decision to alter the procedures for using secondary information sources for completing screeners. Screeners are often closed with no household contact in many cases where no household members would be selected. For example, in many dwelling units, to meet NHANES sampling requirements, white, non-low-income household members are not selected. After several unsuccessful contact attempts (not in the case where a household member has actively refused), an apartment manager or other outside source could be used to determine that no household members would be selected, and the screener could be closed. Using data from 2017, without this procedure the screener response rate would have dropped from 93.1% to 80.1%.

Non-Response Bias Module at the Screening Stage

The addition of questions to provide additional information about for non-response bias analysis is critical for the study, but it adds increases burden on the participants in terms of both time and provision of personal information they may be reluctant to share if they do not want to participate in the full study. The proposed incentive test will help us determine if we need an incentive to maintain historically high screener rates, even with the addition of the non-response bias survey.

Information collected in the screener and non-response bias survey can be used to adjust the weights for respondents to the NHANES interview and examination. Weighting is only effective in reducing bias to the extent that the screener variables used in weighting are related to survey estimates. Therefore, to improve non-response adjustments, a non-response bias module was added to the screener for 2019 (see **Appendix A**). The information collected from this module will be used to explore different weight adjustments to reduce non-response bias.

To improve screener response rates, and consequently increase response to the non-response bias module, we would like to test the use of prepaid or promised incentives. Prepaid incentives will

be included in the advance letter (see a draft of the advance letter in **Appendix B**). The purpose of the prepaid incentive is to encourage review of the advance letter and improve recognition of NHANES so the interviewer is not the first to introduce NHANES. Small token prepaid incentives also increase survey participation by encouraging respondents to reciprocate the survey’s advance token^{3,4}. Promised incentives will be paid by the interviewer once a household member completes the non-response bias module at the screening stage. The promised incentive will be mentioned in the advance letter and by the interviewer when making contact with the sampled household. Promised incentives directly encourage participation, as receipt is contingent on completing the screener. Promised incentives have been shown to be effective for face-to-face surveys^{5,6}. A number of health related face-to-face surveys have tested and currently use incentives, such as the National Survey of Family Growth and the Medical Expenditure Panel Survey^{7,8}.

To maximize response to the nonresponse bias screener module, the pilot test plans to test a prepaid incentive of \$2 and a promised incentive of \$5. The pilot will include a control group that will not receive a prepaid or promised incentive. This results in three experimental groups, \$0, \$2 prepaid only, and \$5 promised only. There is no condition that will receive both incentives.

The use of incentives at the screener component is expected to have a modest effect on response rates (see **Appendix C** for detectable effects), but is expected to reduce the number of contacts required for a completed SP interview. The savings of effort can then be used to increase effort, or contact attempts, for the SP or MEC components to improve response in those later components. Table 1 below shows the potential cost savings with a reduction in effort of 0.5 contacts per complete and increase in response of 3.1% – the minimum detectable difference with 4 stands. We use the sample size for 2017 and proposed prepaid and promised incentive amounts, but assume fixed results in terms of response increase and effort reduction for the various incentive experimental groups for illustration.

Table 1. Cost savings assumptions for experimental incentive conditions

Factors	NHANES 2017	\$2 Pre	\$5 Promised
---------	-------------	---------	--------------

Prepaid Amount per case	\$0	\$2	\$0
Prepaid Cost (total)	\$0	\$27,668	\$0
Promised Amount per case	\$0	\$0	\$5
Promised Cost (total)	\$0	\$0	\$66,675
Contacts/Complete	4.1	3.6	3.6
Cost/Contact	\$30	\$30	\$30
Sample (2017)	13,834	13,834	13,834
Response Rate	93.1%	96.2%	96.2%
Number Completed	12,897	13,308	13,308
Cost per Complete	\$123	\$110	\$113
Total Savings	--	\$172,952	\$133,080

The literature on incentives is fairly conclusive: incentives are effective at increasing response, and prepaid incentives are more effective than promised. However, it should be noted that much of this work is in the context of mail and telephone surveys. For this pilot, we will test the effectiveness of prepaid vs promised incentives, taking into consideration improvements in data quality, improvements in response, reductions in effort, and resulting cost.

We will use a decision matrix to evaluate the effectiveness of the incentives at this stage, with improvements in data quality being the most important evaluation criterion. For example, if indications of bias are reduced in the primary outcome measures discussed at the end of this section below, this will indicate a positive effect from the incentive even if we observe no significant differences, or small positive differences, in response or level of effort. Since incentives are tested for different amounts at different survey stages, we will consider the magnitude of the difference between incentive levels to determine the optimal amount. Alternatively, the influences of bias may be unaffected by the incentive, potentially resulting in no change to our outcome measures discussed below. However, in this case, if we observe significant increases in response (or MEC yield) this suggests improved study representativeness since nonresponse is decreased. We will thoroughly evaluate the outcomes of the incentive experiment in a report before making recommendation for continued use.

The non-response bias module incentive test will address the following research questions:

1. Impact on data quality and non-response bias:

- Does the use of incentives affect respondent compositions, or result in biases in key health measures?
- 2. Impact on response rates:**
- Does the use of the incentives improve response rates to the screener and the non-response bias module?
- 3. Impact on level of effort and cost effectiveness:**
- Does the use of incentive decrease the level of effort?
 - What is the most cost effective amount, when yield (or reduction in effort) and cost are considered?

The measures that will be used to compare these are:

1. Impact on data quality and non-response bias:

Effects on Household Screener: A goal of increasing response is to bring in groups that are under-represented, or that have different health characteristics. Differences in responses to the non-response bias module questions, both overall and by demographic groups, could show a reduction in bias with increased incentives. For example, are “healthier” people more likely to respond to the screener with an incentive?

To answer this, pairwise comparisons will be made between the no-incentive group, the prepaid incentive group, and the promised incentive group. These will be controlled for multiple comparisons using a Bonferroni or other adjustment (see **Appendix C**). Since the NHANES sample is highly clustered in only a few PSUs, we will control for area effects through the sample design (Appendix D).

For non-response bias module question 1, general health is asked of the household respondent, on a scale of 1 (excellent) to 5 (poor). If the average response for the prepaid incentive group is significantly different from the average response for the no-incentive group, then that would indicate the incentive may be increasing response rates among different types of people. Having different types of people who would not have responded without the incentive would indicate that the incentive improves the data quality of the survey to the extent that outcome statistics are correlated with the self-reporting general health question.

Pairwise comparisons for the other module questions will also be made and tested for significance. For example, the percentage of diagnosed diabetics in the no-incentive group will be compared to the percentage of diagnosed diabetics in the promised incentive group. Any indication that these groups are different would suggest that the incentive has an effect on the types of people who respond to the screener.

Differences in the representations of different demographic groups (e.g., gender, age, race, ethnicity, or income) is not a direct indicator of data quality, but it does show that the incentive is changing respondent compositions (that is, not bringing in more of the same respondent type). As above, we will make pairwise comparisons between the no-incentive group, the prepaid incentive group, and the promised incentive group. For example, the percentage of Asians in the no-incentive group will be compared to the percentage of Asians in the prepaid incentive group. Again, significant differences would indicate that different types of people are being included with the incentives, which could be used to show that data quality is improved to the extent that outcome statistics are correlated with the demographic variables.

In addition to the percentage of Asians, we will also compare the percentage of Hispanics, the percentage of blacks, the percentage of males and females, the percentage of age groups (20-29, 30-39, 40-59, 60+), and the percentage of screeners that have an SP screen in as low income. It is possible that we see no change in compositions, but are more successful in identifying dwelling units with no SPs (respondent composition does not meet sampling criteria). Conditions that result in more resolved eligibility statuses are also of value.

Effects on SP Interview and MEC Exam: We will examine means and proportions for selected key statistics from the SP interview and exam, including the following key survey outcomes for this study:

- Obesity prevalence
- Hypertension prevalence
- High cholesterol prevalence
- Diagnosed diabetes prevalence

- Mean of birth weight

These will be examined by screener-stage incentive condition (i.e. non-response bias module incentive), by demographic subgroup, and accounting for the SP interview incentive conditions. As before, pairwise comparisons will be used in this analysis. For example, the obesity prevalence for SPs from households with a prepaid screener incentive will be compared to obesity prevalence for SPs from households with no screener incentive. If we know from the previous analysis that the demographic characteristics or health characteristics (from the non-response bias module) of the no-incentive group are different from the prepaid incentive group, then we would expect some differences in the outcomes as well. This would clearly show that the incentives reduce bias.

2. Impact on response rates:

The current 2018 response rate for the screener is 88.9 covering ten stands in 2018. We project an increase of approximately 4 percentage points would be a significant effect. Our proposal includes three monetary conditions of increasing amounts. For larger amounts we would expect a difference of 3 percentage points from conditions with lower amounts, or observe non-significant differences between the lower amounts and the control condition of no incentive (with corresponding significant difference between the larger amount and no incentive). **Appendix C** shows that a difference can be detected within 6 stands yielding a minimum detectable difference of 2.6 percentage points or 2.3 percentage points for 8 stands.

3. Impact on level of effort and cost effectiveness:

We expect the screener incentive to offer reductions in effort defined by average contact attempts per complete. Combined with increases in response, complimentary reductions in effort will result in substantial cost savings compared to a no incentive condition. We project a reduction of approximately 0.5 contact attempts per complete. Table 1 shows that all incentive conditions result in reduced cost with a reduction of this magnitude and a response rate increase of 3.1 percentage points (below our projected increase).

SP In-Home Interview Incentives

Sample Person (SP) interview response has shown the most dramatic decline, accounting for much of the decline in the overall rate (Screener * SP in-home interview * MEC). Declines in SP in-home interview response began in 2011 and have more recently accelerated. In 2011, the conditional SP in-home interview response rate was 75.1% compared to 57.4% in 2017. Clearly, this is an area of great potential for reducing bias, improving response, and increasing the number of MEC interviews.

We expect incentives to have a direct effect on SP in-home interview response. This is based on conclusions from the incentive literature that have shown that incentives are most effective when response is low⁴ and burden is high⁶, factors that are true for the SP in-home interview. The SP in-home interview can take up to one hour and may ask for information that respondents find sensitive.

We propose to test the use of a promised incentive contingent on completing the SP in-home interview. Often the MEC incentive is used at the interview component to elicit cooperation for both the SP in-home interview and MEC examination. This has the unintended consequence of perceptually increasing the burden covered by the token incentive, reducing the overall effectiveness of it.

The promised SP in-home interview incentive would test three incentive levels. The levels are \$0, \$20, and \$40. The purpose of the three conditions is to determine the most effective level at motivating response compared to a no incentive condition, and to determine if a “gradient” of impact on data quality and bias exists. The two monetary levels were chosen based on the experimental results of other Federal studies. We offer examples of current Federal survey incentives below:

- National Survey on Drug Use and Health (NSDUH) - \$30 since 2001.
- Medical Expenditure Panel Survey (MEPS) - \$50 since 2011, increased from \$30 and has tested up to \$70. This survey requires the participant to gather medical bills and insurance receipts in order to respond.

- National Survey of Family Growth (NSFG) - \$40 since 2006 (with a increased incentive for refusal conversion). A recent experiment found that increasing the base incentive to \$60 did not reduce non-response bias.

We hypothesize that the \$40 incentive level will demonstrate larger and significant increases in response compared to \$20 or no incentive conditions. This will be evaluated empirically, but our hypothesis is based on current incentive levels used on other studies, all of which are based on experiments with different incentive levels. We include the \$20 level since the incentive will be offered in the context of other incentives at the screener and MEC examination. It is possible that when considering cost, given the magnitude of the increase in response due to the incentive, the \$20 amount offers the greatest return. For example, MEPS tested \$70 and NSFG tested \$60. While both MEPS and NSFG showed increases in response from the higher incentive amount, neither found the higher amounts to improve the quality of the data significant or be cost effective. We speculate a reason for this is that incentive saliency is linked to common monetary denominations and therefore suggest amounts mimicking these (for example \$20 and \$40).

It is important to note that screener and SP in-home interview incentives are tested concurrently. The power analysis in **Appendix C** shows power to detect a difference in up to 8 stands. Testing screener incentives and SP in-home interview incentives separately would require a substantial investment in time, possibly delaying other pilot tests. We believe both the non-response bias module and SP in-home interview incentives can be tested concurrently and show potential assignments at the segment level in **Appendix D**, combining both and ensuring diversity across segment compositions.

The SP in-home interview incentive test will address the following research questions:

1. **Impact on data quality and non-response bias:**
 - Do promised incentives affect respondent compositions, or reduce biases in key interview measures?
2. **Impact on response rates:**

- Do promised incentives yield increases in response? What is the relative magnitude on response of different incentive amounts? What effect does increased response (due to an incentive) have on MEC participation?

2. Impact on level of effort and cost effectiveness:

- What is the most cost effective amount, when return (increase in response) and cost are considered?

The measures that will be used to compare these are:

1. Impact on data quality and non-response bias:

SP Interview: As noted earlier, decreases in SP interview participation are disproportionate by demographic subgroups. Similar to the screener-stage incentive analysis, we will make pairwise comparisons between the no-incentive group, the \$20 incentive group, and the \$40 incentive group. For example, the percentage of Asians in the no-incentive group will be compared to the percentage of Asians in the \$40 incentive group. Again, significant differences would indicate that different types of people respond when they are promised interview incentives, which suggested that data quality is improved. In addition to the percentage of Asians, we will also compare the percentage of Hispanics, the percentage of blacks, the percentage of age groups (20-29, 30-39, 40-59, 60+), and the percentage of SPs who screen in as low income.

We hypothesize that SPs who are promised an interview incentive will differ in key health measures from SPs who are not promised an interview incentive. We will examine means and proportions for selected key statistics from the SP interview (including diagnosed diabetes rate and mean of birth weight, which are considered key survey outcomes for this study). The proposed screener-stage incentives may affect these key statistics, so that will be considered.

For example, the average birth weight for SPs with a \$40 incentive will be compared to the average birth weight for SPs with no interview incentive. If we know from the previous analysis that the demographic characteristics of the no-incentive group are different from the \$40 incentive group, then we would expect some differences in the outcomes as well. This would suggest that the incentives reduce bias.

MEC Examination: The MEC exam is the most important component of NHANES. MEC exam incentives are unchanged to assess the effect of screener and SP interview incentives. Any changes in MEC yield are expected to have effects on key statistics. Increases in SP interview participation may affect individual component participation in the MEC exam. We will examine key statistics from the MEC exam (including obesity prevalence, hypertension prevalence, and high cholesterol prevalence, which are considered key survey outcomes for this study) and evaluate participation rates in individual MEC components by incentive group. If MEC yield positively increases with a significant increase from the SP incentive, but participation in key component measures have a lower net yield, this may not support adopting the incentive.

In addition, MEC nonrespondents who completed the interview can be grouped with MEC respondents based on their interview responses, which would decrease potential nonresponse bias. Currently, SPs who do not want to travel to the MEC (for any reason) have no incentive to complete the interview. Very few SPs complete the interview and do not have an examination at the MEC, meaning very little is known about the MEC nonrespondents. To reduce nonresponse bias through weighting adjustments, more information about nonrespondents is needed.

2. Impact on response rates:

Response (SP interview): Our proposal includes two monetary conditions of increasing amounts. For the larger amount (\$40). We should be able to detect a difference of 6 percentage points from the lower amount condition of \$20, or observe non-significant differences between the lower amount and the control condition of no incentive (with corresponding significant difference between the larger amount and no incentive). **Appendix C** shows that a difference can be detected within 6 stands yielding a minimum detectable difference of 5.9 percentage points, or 5.1 percentage points for 8 stands.

Response rates will also be examined by demographic characteristics within each incentive group, since demographics (e.g., gender, age, race, ethnicity, or income) are known when SPs are identified. A significant relationship between response status and an auxiliary variable could indicate potential bias in the health statistics, to the extent that the auxiliary characteristic and the

outcome of interest are correlated. SPs who respond to the interview will be compared to interview nonrespondents. For example, the difference between the response rate of male SPs who receive a \$40 incentive and the response rate of male SPs who receive no interview incentive will be tested for significance. Response rates will be calculated with screener base weights to account for the different sampling rates by domain.

Response and Net Yield (MEC exam): We will examine the effect changes in SP interview response have on MEC participation and net yield. The current conditional MEC response rate is 93.3% for 2018. This shows the importance of increasing participation in the SP interview to meet the goal of 5,000 completed yearly MEC exams (increasing the SP interview rate provides the greatest opportunity for increasing overall response). Many who are refusing the SP interview may be doing so due to reluctance to complete the MEC exam. Therefore, increases in the SP interview rate may result in a decrease in the conditional MEC exam rate. Net yield for MEC exam participation will provide a more accurate measure. This is because disproportionate changes in response for the SP interview and MEC exam may still result in more completed MEC examinations. For example, using number of completed SP interviews for 2017; a 6 percent increase in SP response with an 8 percent decrease in MEC examinations results in a net increase of 49 additional MEC examinations. Significant SP interview increases that result in negative MEC exam yields will not support adopting the incentive level found significant.

3. Impact on level of effort and cost effectiveness:

We expect the interview incentive to offer reductions in effort defined by average contact attempts per complete. Combined with increases in response, complimentary reductions in effort will result in substantial cost savings compared to a no incentive condition. We project a reduction of approximately 0.5 contact attempts per complete. This is within the detectable difference for 6 stands of 0.31, or 0.27 for 8 stands. Our estimate for the reduction in contacts per complete is conservative. This is because for many SPs the SP interview is their first contact, where unlike the screener, the interview request is far more substantial. We expect the reduction in contacts to come from SPs who were also the screener respondent, or household members who are responsible for SP children or are primary decision makers for a selected family.

Power Analysis for Incentive Experiments

A power analysis to show the minimum detectable differences in response rate and contact attempts, for both the screener and interview, is shown in **Appendix C**. A minimum of 6 to 8 stands is recommended to find a detectable difference in the higher incentive groups.

Appendix A

The Non-Response Bias Module

SCQ.600 First, I have some general questions about your health.

Would you say your health in general is . . .

- excellent,..... 1
- very good,..... 2
- good,..... 3
- fair, or..... 4
- poor?..... 5
- REFUSED..... 7
- DON'T KNOW..... 9

SCQ.610 Are you now taking any medications prescribed by a health professional such as a doctor or dentist?

- YES..... 1
- NO..... 2 (SCQ.630)
- REFUSED..... 7 (SCQ.630)
- DON'T KNOW..... 9 (SCQ.630)

SCQ.620 How many prescription medications do you currently use or take? Would you say...

- 1 to 2, 1
- 3 to 5, or..... 2
- 6 or more?..... 3
- REFUSED..... 7
- DON'T KNOW..... 9

SCQ.630 Has a doctor or other health professional ever told you that you had diabetes?

INTERVIEWER INSTRUCTION:
IF DIABETES ONLY DURING PREGNANCY, CODE NO.

- YES..... 1
- NO..... 2
- BORDERLINE OR PREDIABETES 3
- REFUSED..... 7
- DON'T KNOW..... 9

SCQ.640 Has a doctor or other health professional ever told you that you had hypertension (hy-per-ten-shun), also called high blood pressure?

INTERVIEWER INSTRUCTION:

IF HIGH BLOOD PRESSURE ONLY DURING PREGNANCY, CODE NO.

IF RESPONDENT SAYS “HIGH NORMAL BLOOD PRESSURE”, “BORDERLINE HYPERTENSION” OR “PREHYPERTENSION” CODE NO.

YES.....	1
NO.....	2
REFUSED.....	7
DON'T KNOW.....	9

HELP SCREEN:

Hypertension (High Blood Pressure): A repeatedly increased blood pressure with the first number 140 or higher and the second number 90 or higher.

Appendix B

Draft of Proposed Advance Folded Postcard and Letter



U.S. Department of
Health and Human Services
Centers for Disease
Control and Prevention

U.S. Department of Health and Human Services
Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toldeo Road, Room 4322
Hyattsville, MD 20782

First Class Mail
U. S. Postage
PAID
Suburban MD
Permit No. 6379

The Household Residing at
123 Anywhereville Ave.
Anywhere, US 12345



Be a part of something
unique. Be a part
of the nation's
premier health
survey.



The National Health and Nutrition Examination Survey (NHANES) is in your neighborhood and your household has been selected for an important first step.

In the next week you will receive a letter with more information about what to expect. Please watch for this letter. **The letter is important and includes a small cash token of our appreciation for your time.**

Sponsored by the National Center for Health Statistics (part of the CDC), information from NHANES has been used to solve health problems, develop health programs, and improve the quality of health care in the United States.

For more information visit: www.cdc.gov/NHANES

Photo is for illustrative purposes only. Any person depicted in the photo is a model.





NATIONAL CENTER FOR HEALTH STATISTICS

National Health and Nutrition Examination Survey

Resident
Address 1, Address 2
City, ST ZIP

DATE

From the Director,

Your household has been selected to take part in an important national health study. This is the National Health and Nutrition Examination Survey or NHANES. NHANES provides important information about the health and diets of people in the United States. The National Center for Health Statistics, which is a part of the Centers for Disease Control and Prevention (CDC), conducts this important survey. Over the years, this survey has helped us understand the impact of changes in the foods we eat and the health care we receive.

In the next few weeks, one of our study representatives will make a personal visit to your home. They will see if you or a member of your family will be invited to take part in this study. Here is what to expect:

- She or he will show you official Centers for Disease Control (CDC) identification and ask some easy questions about you and other family members.
- [BLANK/You will receive \$5 for answering questions about your household.]
- Answering the questions only takes a few minutes and is voluntary. You may choose not to answer any question. There will be no penalties or loss of benefits of any kind from refusing to answer.
- If you or other household members are chosen, you'll represent thousands of people in the U.S. and your community to help us better understand the health of people like you in the United States.
- If you have questions, please call Dr. Tony Nguyen of my staff toll-free at 1-800-452-6115.

We ask that you take a few minutes to speak with our interviewer when they visit. [BLANK/To thank you in advance for speaking with our interviewer and learning more about NHANES, we are enclosing a \$2 bill. This token of appreciation is yours to keep whether or not you decide to participate.] [BLANK/We will give you a \$5 token of appreciation for answering questions about you and your household.]

Thank you in advance for helping this important study.



Sincerely,

Kathryn S. Porter, MD, MS,
Captain, U.S. Public Health Service
Director, Division of Health and Nutrition Examination Surveys
National Center for Health Statistics, Centers for Disease Control and Prevention

This survey is a Federal program authorized by Section 306 of the Public Health Service Act (42 U.S.C. 242k). All of your answers will be kept in strict confidence according to Section 308(d) of the Public Health Service Act (Title 42, U.S. Code, 242m(d)) and the Confidential Information Protection and Statistical Efficiency Act (Title 5 of PL 107-347). We will use your information only for statistical research and reports. Your answers will be added to others, so no one can identify which are yours.



Centers for Disease
Control and Prevention
National Center for
Health Statistics

Appendix C: Power Analysis

Screener Experiment

Screener Response Rate

For 2017 and the first 10 stands of 2018 (25 total stands), 23,705 DUs were released, and 3,921 of them were coded as vacant or not a DU, leaving 19,784 eligible DUs, or about 790 per stand. Of the eligible DUs in those 25 stands, 18,083 responded to the screener, for a 91.4% screener response rate.

The table below shows for a given number of stands, the estimated number of eligible DUs and the estimated number of DUs that would be assigned to each of the 3 screener experimental groups (\$0, \$2 prepaid, and \$5 promised).

A power analysis was conducted for a comparison of proportions, using significance levels of the standard $\alpha = 0.05$ for a two-tailed test and 80% power. The table shows the minimum detectable difference (MDD) in screener-level response rate between any two of the screener experimental groups, assuming one of the groups is at about 91.4%. For example, response rates for the group with \$0 could be compared to response rates for the group with \$5 promised. We can also compare the screener response rate from the last 25 stands to see if the experimental groups are significantly different. The impact of the clustered design on the variances of differences between conditions is expected to have only a slight increase due to the incentive experiment sample design (see **Appendix D**) in which all conditions occur in the same set of PSUs. In addition, the use of linear models that include auxiliary variables will reduce variances associated with the comparisons. Therefore, these calculations assume a design effect of 1.0. We did not control for multiple comparisons in the power analysis. Controlling for multiple comparisons will be considered whenever more than one comparison is tested as a family of tests. For example, with 3 conditions, there are three tests conducted, and with the Bonferroni adjustment would conservatively reduce the critical α to 0.017 and the MDDs would be larger. It is well known that the Bonferroni adjustment may be too conservative, and so other multiple comparison methods will be considered.

# Stands	# Eligible DUs	# Eligible DUs per	MDD in Screener RR
----------	----------------	--------------------	--------------------

		Screeener Group	Experiment	Between Groups	Any Two
3	2,370	790		0.035	
4	3,160	1,053		0.031	
5	3,950	1,317		0.028	
6	4,740	1,580		0.026	
7	5,530	1,843		0.024	
8	6,320	2,107		0.023	

With six stands in the experiment, the true screener response rate would have to increase about 2.6 percentage points (for example, from 91.4% estimated for the \$0 group to about 94.0% for a higher incentive group) to detect significance. With eight stands, the increase would have to be about 2.3 percentage points to detect significance.

Screeener Contact Attempts

The total number of contact attempts to complete the screener for eligible DUs (including DUs that completed the screener and DUs that did not complete the screener) divided by the total number of completed screeners in these 25 stands was 4.1. This is an estimate of the total level of effort required to complete a screener. The table below shows the minimum detectable difference (MDD) in the mean number of contact attempts between any two of the screener experimental groups. As with response rates, total contact attempts for the group with \$0 could be compared to total contact attempts for the group with \$5 promised, for example.

# Stands	# Eligible DUs	# Eligible DUs per Screener Group	MDD in Screener Contact Attempts Between Any Two Groups
3	2,370	790	0.47
4	3,160	1,053	0.41
5	3,950	1,317	0.37
6	4,740	1,580	0.33
7	5,530	1,843	0.31
8	6,320	2,107	0.29

With six stands in the experiment, the true number of contact attempts to complete a screener would have to decrease from about 4.1 to about 3.77 to detect significance. With eight stands, the decrease would have to be from 4.1 to about 3.81 contact attempts to detect significance.

Screener Cost per Complete

Screener cost per contact will be estimated over the previous 25 stands (currently 2017 and the first 10 stands of 2018). This amount is estimated since interviewers do not charge their time separately for various interview activities in a stand (e.g, screener attempt or SP in-home interview contact attempt). The previous 25 stands are used for consistency with response rate and contacts per complete measures. This also provides stability since cost may vary by stand due to factors such as population density, proportion of locked structures, racial and ethnic compositions and other factors. This amount is used in the cost formula below to determine the average cost per complete for each of the 3 incentive experimental groups.

The average cost per complete is determined by adding the incentive cost and the product of average cost per contact, number of screener completes, and average number of contacts. This sum is divided by the total number of screener completes and is separately calculated for each of the 3 incentive groups.

- Number of cases in experimental incentive condition (n_a)

- Number of cases completing the in-person screener in experimental incentive condition (n_b)
- Total prepaid incentive cost (I_{pre}) = prepaid incentive amount * sample in experimental group (n_a)
- Total promised incentive cost (I_{post}) = promised incentive amount * number of screeners completed in experimental group (n_b)
- Cost per contact (C_a) = average cost per contact (average from prior 25 stands)
- Contacts per complete (V_a) = average number contacts per complete for experimental group

$$\text{Cost per Complete} = [I_{pre} + I_{post} + (C_a * n_b * V_a)] / n_b$$

Cost per complete is largely influenced by the screener response rate (yielding number of screener completes – n_b) and average contacts per complete (V_a). Increases in screener response are not necessary to see a reduction in cost per complete, but decreases in the average contacts per complete are necessary.

Cost per complete will be calculated across all stands in the experiment for each experimental condition. This will be used to assess the effective cost of the incentives provided.

SP In-Home Interview Experiment

SP In-home Interview Response Rate

For 2017 and the first 10 stands of 2018 (25 total stands), 13,577 SPs were identified, or about 543 per stand. Of those identified SPs, 7,639 responded to the interview, for a 56.3% interview response rate.

The table below shows for a given number of stands, the estimated number of eligible SPs and the number of SPs that would be assigned to each of the 3 interview experimental groups (\$0, \$20, \$40). In practice, all SPs in the same DU would be included in the same interview experimental group.

A power analysis was conducted for a comparison of proportions, using significance levels of the standard $\alpha = 0.05$ for a two-tailed test and 80% power. The table shows the minimum detectable difference in interview-level response rate between any two of the screener experimental groups, assuming one of the groups is at about 56.3%. For example, response rates for the group with a \$0 incentive could be compared to response rates for the group with a \$40 incentive. We can also compare the screener response rate from the last 25 stands to see if the experimental groups are significantly different. (These calculations also assume no design effect and we did not control for multiple comparisons in the power analysis.)

# Stands	# Eligible SPs	# Eligible SPs per Interview Experiment Group	MDD in Interview RR Between Any Two Groups
3	1,629	543	0.083
4	2,172	724	0.072
5	2,715	905	0.065
6	3,258	1,086	0.059
7	3,801	1,267	0.055
8	4,344	1,448	0.051

With six stands in the experiment, the true interview response rate would have to increase about 6 percentage points (for example, from 56.3% estimated for the \$0 group to 62.2% for a higher incentive group) to detect significance. With eight stands, the increase would have to be about 5 percentage points to detect significance.

SP In-Home Interview Contact Attempts

The total number of contact attempts to complete an interview for SPs (including SPs that completed the interview and SPs that did not complete the interview) divided by the total number of completed interviews in these 25 stands was 8.0. This is an estimate of the total level of effort required to complete an interview. The table below shows the minimum detectable difference in the mean number of contact attempts between any two of the interview experimental groups. As

with response rates, total contact attempts for the group with a \$20 incentive could be compared to total contact attempts for the group with a \$40 incentive, for example.

# Stands	# Eligible SPs	# Eligible SPs per Interview Experiment Group	MDD in Interview Contact Attempts Between Any Two Groups
3	1,629	543	0.44
4	2,172	724	0.38
5	2,715	905	0.34
6	3,258	1,086	0.31
7	3,801	1,267	0.29
8	4,344	1,448	0.27

With six stands in the experiment, the true number of contact attempts to complete an interview would have to decrease from about 8.0 to about 7.69 to detect significance. With eight stands, the decrease would have to be from 8.0 to about 7.73 contact attempts to detect significance.

Interview Cost per Complete

Interview cost per contact will be estimated over the previous 25 stands (currently 2017 and the first 10 stands of 2018). This amount is used in the cost formula below to determine the average cost per SP in-home interview for each of the 3 incentive experimental groups.

The average cost per SP in-home interview is determined by adding the total promised incentive cost, and the product of average cost per contact, number of SP in-home interviews, and average number of contacts. This sum is divided by the total number of SP in-home interviews and is separately calculated for each of the 3 incentive groups.

- Number of completed SP in-home interviews in incentive condition (n_b)
- Number of cases completing the SP in-home interview in experimental incentive condition (n_b)

- Total promised incentive cost (I_{post}) = promised incentive amount * number of SP in-home interviews completed in experimental group (n_b)
- Cost per contact (C_a) = average cost per contact (average from prior 25 stands)
- Contacts per complete (V_a) = average number contacts per SP in-home interview for experimental group

$$\text{Cost per Complete} = [I_{\text{post}} + (C_a * n_b * V_a)] / n_b$$

Cost per SP in-home interview is influenced by the average contacts per complete (V_a). Increases in interview response will not affect the cost per complete since the incentive is only paid for completed interviews and is a fixed cost within each experimental condition. Decreases in the average contacts per complete are necessary for any decreases in cost.

Cost per SP in-home interview will be calculated across all stands in the experiment for each experimental condition. This will be used to assess the effective cost of the incentives provided.

Analysis of Clustered Data

To account for the clustered design of this experiment, generalized estimating equations (GEEs) methodology will be used to analyze correlated data on the segment level. STATA will be used with an xtgee command.

Appendix D

Incentive Test Protocol

Screener Incentive – 3 levels, \$0, \$2 prepaid, \$5 promised

SP Interview Incentive – 3 levels, \$0, \$20, \$40

Crossing all of these cells gives $3 \times 3 = 9$ different cells

Each primary sampling unit (PSU), or stand location, is comprised of 24 segments. With nine different cells for testing, two or three of the 24 segments in each survey location will be assigned to a unique treatment cell. This will ensure that each segment would get the same set of incentives, say at the screener \$2 prepaid, and \$20 after the SP in-home interview. This would reduce potential mistakes by the field interviewer, who would know an entire segment would have the same incentives, rather than having potentially different incentives at every household. This would also minimize the potential problem of neighbors knowing they received different incentives, which could affect the incentive experiment. If neighbors in the same segment talk to each other, they would be getting the same incentives, and all incentives would be paid at the household (screener and/or interview).

SPs from different segments would still get the same incentive for MEC participation, but they could talk to each other about the different screener or interview incentives if they are at the MEC together or if they live in different neighborhoods but know each other. We anticipate this to be an unlikely occurrence. NCHS will provide guidelines for responding to concerns from respondents about this issue and share them with field staff.

The current procedure to create segments results in 24 segments (for most survey locations) that are sorted from lowest minority density to highest, so segment 1 always has the highest percentage of white/other population, and segment 24 has the lowest. Hispanic and black percentages of the population tend to be highest in the higher-numbered segments.

Median income for a segment is strongly correlated with segment order as well. Over the past 20 survey locations, segments 1-6 have an average median income of \$73,712, for segments 7-12 it is \$64,354, for segments 13-18 it is \$54,934, and for segments 19-24, it is \$43,683.

Because of the demographic differences in the segments, the assignment of incentives will not be completely random to distribute incentive assignments across the different groups. This will prevent, for example, segments 1-6 (the higher income segments) from all getting the \$40 interview incentive, or all getting the \$0 interview incentive.

The protocol for assigning incentives to segments will be as follows: The 3 cells defining all possibilities of screener prepaid incentive and screener promised incentive will randomly be notated with letters A-C. For example, the \$2 prepaid might be “A” and the \$5 promised might be “B”, etc. Similarly, the 3 interview incentive cells will randomly be notated with letters X-Z.

The table below shows how the incentives will be assigned to each segment in the first eight survey locations of the study. For example, if \$2 prepaid is “A” and \$20 interview is “X”, then those will be the incentives for segment 1. Note that all three-segment blocks (1-3, 4-6, etc.) have all of the possible screener and interview incentives. In this way, the incentive allocation will be balanced among different types of segments, which will mitigate the clustering within segments. The “AX” incentives will be in segments 1, 4 and 19 in survey location 1, in segments 9, 12 and 24 in survey location 2, in segments 2 and 11 in survey location 3, etc.

Segment	Survey Location 1	Survey Location 2	Survey Location 3	Survey Location 4	Survey Location 5	Survey Location 6	Survey Location 7	Survey Location 8
1	AX	BX	CZ	AZ	BY	CY	AY	BZ
2	BY	CZ	AX	CY	AZ	BX	BZ	CX
3	CZ	AY	BY	BX	CX	AZ	CX	AY
4	AX	BZ	CZ	AY	BY	CX	AZ	BX
5	BZ	CX	AY	CZ	AX	BY	BX	CY
6	CY	AY	BX	BX	CZ	AZ	CY	AZ
7	AZ	BZ	CY	AY	BX	CX	AX	BY
8	BX	CY	AZ	CX	AY	BZ	BY	CZ
9	CY	AX	BX	BZ	CZ	AY	CZ	AX

10	AZ	BY	CY	AX	BX	CZ	AY	BZ
11	BY	CZ	AX	CY	AZ	BX	BZ	CX
12	CX	AX	BZ	BZ	CY	AY	CX	AY
13	AY	BY	CX	AX	BZ	CZ	AZ	BX
14	BZ	CX	AY	CZ	AX	BY	BX	CY
15	CX	AZ	BZ	BY	CY	AX	CY	AZ
16	AY	BX	CX	AZ	BZ	CY	AX	BY
17	BX	CY	AZ	CX	AY	BZ	BY	CZ
18	CZ	AZ	BY	BY	CX	AX	CZ	AX
19	AX	BX	CX	AZ	BZ	CZ	AY	BY
20	BY	CY	AY	CY	AY	BY	BX	CX
21	CZ	AZ	BZ	BX	CX	AX	CZ	AZ
22	AY	BY	CY	AX	BX	CX	AZ	BZ
23	BZ	CZ	AZ	CZ	AZ	BZ	BY	CY
24	CX	AX	BX	BY	CY	AY	CX	AX

References

1. 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(2):186-211.
2. Parker E, Field J. Demographic Survey Overview: Census Scientific Advisory Committee September 17-18. 2015; <https://www2.census.gov/cac/sac/meetings/2015-09/2015-Parker.pdf>.
3. Dillman DA, Smyth JD, Christian LM. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons; 2008.
4. Singer E, Ye C. The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*. 2013;645(1):112-141.
5. Mercer A, Caporaso A, Cantor D, Townsend R. How Much Gets You How Much? Monetary Incentives and Response Rates in Household Surveys. *Public Opinion Quarterly*. 2015;79(1):105-129.
6. Singer E, Van Hoewyk J, Gebler N, McGonagle K. The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics*. 1999;15(2):217.
7. Lepkowski JM, Mosher WD, Groves RM, West BT, Wagner J, Gu H. Responsive Design, Weighting, and Variance Estimation in the 2006-2010 National Survey of Family Growth. *Vital and health statistics Series 2, Data evaluation and methods research*. 2013(158):1-52.
8. Agency for Healthcare Research and Quality. Respondent Payment Experiment with MEPS Panel 13. 2010; http://meps.ahrq.gov/mepsweb/data_files/publications/rpe_report/rpe_report_2010.shtml. Accessed June, 2018.