

Attachment IE-1

The National Survey on Drug Use and Health (NSDUH) Screening and Interview Incentive Experimental Design and Power Analysis

1. Overview

Nonresponse has been increasing in face-to-face surveys in the United States (Williams and Brick, 2018). This trend has also been observed with the National Survey on Drug Use and Health (NSDUH) where nonresponse has been increasing despite the increased level of effort for data collection such as increasing contact attempts. The NSDUH's overall response rate is a product of the screening response rate (SRR) and the main interview response rate (IRR). While both response rate components are decreasing, screening nonresponse has been increasing at a faster rate than interview nonresponse in recent years. NSDUH weighted SRRs dropped to 73.3 percent for the 2018 survey year, compared to 88.0 percent in 2010.

Nonresponse bias in NSDUH estimates can result from nonresponse at the screening phase, the interview phase, or both. One of the approaches Substance Abuse and Mental Health Services Administration (SAMHSA) is considering to stem this decline is testing the impact of a \$5 screening incentive. No incentive is currently offered for completing the screening interview.

Since 2002, the NSDUH has offered a \$30 incentive for completing the main interview to sample members selected after completion of the household screening. A second approach SAMHSA plans to test is increasing the \$30 main interview incentive to \$50. Combined with the \$5 screening incentive, a 2 by 2 experiment is created by crossing the two screening incentive amounts and the two interview incentive amounts.

Multiple studies have shown that incentives tend to increase participation among sample members who are less interested in or involved with the survey topic (Groves, Singer, & Corning, 2000; Groves, Presser, & Dipko, 2004; Groves et al., 2006). Adding a screening incentive and increasing the main interview incentive could increase participation at both the screening and interviewing stages and therefore, reduce the potential for nonresponse bias in key NSDUH estimates from households whose residents are less interested in substance use or mental health issues. The screening and interview incentive experiment will be used to primarily assess:

- 1) The impact of offering a screening incentive on screening response rates (SRR);
- 2) The impact of a higher interview incentive on interview response rates (IRR);
- 3) The impact of offering a screening incentive on data quality in terms of reducing nonresponse bias through the demographic composition of households screened.¹ and
- 4) The impact of the combination of screening and interview incentive amounts on SRRs, IRRs, and the demographic composition of households screened.

Secondarily, the screening and interview incentive experiment will also be used to assess:

- 1) The impact of a higher interview incentive on screening response rates (SRR);¹
- 2) The impact of a higher interview incentive on data quality through reducing nonresponse bias in key estimates from the interviewing phase;

¹ Given that sampled NSDUH households are alerted to the interview incentive in an advance letter and this letter is provided by FIs in person to household members who indicate they have not seen the letter, the interview incentive amount could influence the propensity to complete the screening to determine if anyone is selected for an interview.

- 3) The impact of the combination of screening and increased interview incentive amounts on data quality through reducing nonresponse bias in key estimates from the interviewing phase
- 4) The impact of the screening incentive on the number of contact attempts per completed screening; and
- 5) The impact of the screening incentive or increased interview incentive on the number of contact attempts per completed interview.

SAMHSA would like to evaluate whether adding a screening incentive and/or increasing the interview incentive boosts the likelihood of participation in the combined household screening interview and subsequent main interview(s). The current plan is to test this 2 by 2 incentive experiment in a future NSDUH survey. This appendix includes a description of the incentives experiment and the power calculations made for assessing nonresponse bias at the screening and interview phases.

2. Experimental Conditions for the Combined Screening and Interview Incentive Experiment

Based on discussions between OMB and SAMHSA about testing incentives, the NSDUH incentive experiment will involve testing both a screening incentive (\$0 versus \$5) and an interview incentive (\$30 versus \$50). The \$5 screening incentive and the interview incentives will be promised incentives. That is, field interviewers (FIs) will only provide the screening incentive to household members who are determined to be eligible screening respondents (SRs)² and FIs will provide the interview incentive after selected respondents complete the interview. The literature on incentives clearly demonstrates monetary incentives are effective in increasing survey cooperation rates. In addition, prepaid incentives have been shown to be more effective than promised incentives (Singer and Ye, 2013). However, much of this research on prepaid incentives has been conducted in the context of mail and telephone surveys. Offering promised incentives for in-person surveys like the NSDUH can function similarly to prepaid incentives in self-administered survey modes. FIs can clearly demonstrate to potential SRs that they have the promised screening and interview incentives and will provide these when stated in the NSDUH materials and scripts. Table 1 presents the four possible combinations of screening and interview incentive amounts that would comprise the four experimental conditions.

Table 1. Four Possible Experimental Conditions for the Combined Experiment

		Screening Incentive Amounts			
		\$0 screening + \$30 interview		\$5 screening + \$30 interview	
Interview Incentive Amounts	1	\$0 screening + \$30 interview	2	\$5 screening + \$30 interview	
	3	\$0 screening + \$50 interview	4	\$5 screening + \$50 interview	

Some important components of the experimental design to note are:

² For household members to be eligible as a screening respondent, the person must be aged 18 years or older and be confirmed as a resident of the household.

- Four segments will be sampled within each state sampling region (SSR);
- Each segment selected will be assigned to one of the four experimental conditions in Table 1, with only one segment assigned to each of the four conditions within an SSR;
- An equal number of segments will be assigned to each of the four experimental conditions across the sample;
- Two FIs will be assigned to each segment/condition within each SSR, so that each interviewer works four segments with each of the four experimental conditions; and
- FIs will be trained and monitored to ensure their work is balanced across the four segments and conditions throughout the entire field period.

This experimental design is intended to:

- 1) minimize the potential for “spillover” errors or effects by assigning each FI four segments representing the four different experimental incentive conditions. Each FI will know the experimental incentive condition for all SDUs within each segment, to reduce the potential for offering an SDU in the same segment the wrong incentive conditions; and
- 2) minimize the need to account for interviewer effects on the screening and interviewing response rates by experimental condition. By assigning each FI four segments representing the four different experimental incentive conditions and ensuring FIs balance their work across the four segments, any interviewer effects should be spread evenly across the four conditions in each SSR.

3. Screening Response Rates and a Screening Incentive

The weighted NSDUH SRR for the most recent completed year (2018) was 73.3%. This is lower than previous years – about 2 percentage points lower than 2017 and over 4 percentage points lower than 2016. Despite increases in the level of effort to complete interviews (as measured in terms of number of contact attempts), recent literature reviews have shown declining response rates (Williams and Brick, 2018). Recent NSDUH experience has been consistent with this trend for SRRs. The most important impact of declining SRRs is the potential impact of screening nonresponse on bias, from both the composition of households screened and the estimates generated from interviews completed in these households. No screening incentive is currently offered to NSDUH screening respondents (SRs). These factors directly informed SAMHSA’s plan to test a promised \$5 screening incentive.

The experimental test of a \$5 screening incentive will be used to assess:

- 1) The impact of offering a screening incentive on SRRs; and
- 2) The impact of offering a screening incentive on data quality in terms of the quality of screening data and on the demographic composition of households screened.

4. Interview Incentives

The weighted NSDUH IRR for the most recent completed year (2018) was 66.6%. This is lower than previous years – about 1 percentage point lower than 2017 and about 2 percentage points lower than 2016. Despite increases in the level of effort to complete screenings and interviews (as measured in terms of number of contact attempts), recent literature reviews have shown declining response rates (Williams and Brick, 2018). Recent NSDUH experience has been consistent with this trend for IRRs.

Since 2002, the interview incentive promised to NSDUH respondents has been \$30. SAMHSA would like to test a higher interview incentive amount of \$50 against the current interview incentive amount. The higher interview incentive amount is proposed based on a review of other nationally-representative in-person surveys that have recently conducted experimental tests to determine the impact of increasing the interview incentive amount. Examples of other nationally-representative in-person surveys reviewed included:

- *Medical Expenditure Panel Survey (MEPS)*. The interview incentive was increased from \$30 to \$50 in 2011. A higher incentive amount of \$70 has been tested, but not implemented.
- *Panel Study of Income Dynamics (PSID)*. Between 2002 and 2015, the PSID interview incentive increased by \$5 increments three times about every 3 to 4 years, from a starting incentive of \$55 in 2003 to the current incentive of \$70 in 2015.
- *National Survey of Family Growth (NSFG)*. The interview incentive has been \$40 since 2006, with an increased incentive promised for nonresponse follow-up. A higher incentive amount of \$60 has been tested, but not implemented.

These examples illustrate the value of testing different interview incentive amounts to determine whether increased and/or additional incentives reduce nonresponse bias in key estimates. Consistent with this approach, SAMHSA would like to experimentally test the current \$30 versus a \$50 interview incentive. This test of two interview incentive amounts will be used to assess:

- 1) The impact of a higher interview incentive on SRRs and IRRs;¹
- 2) The impact of a higher interview incentive on data quality through reducing non-response bias in key estimates, either from the screening or interviewing phase; and
- 3) The impact of the combination of screening and increased interview incentive amounts on data quality through reducing non-response bias in key estimates from the interviewing phase.

5. Power Analysis

The sample will involve 89 state sampling regions (SSRs), an estimated 356 segments, and 12,774 sample dwelling units (SDUs) to yield approximately 8,110 completed screening interviews and 4,000 completed interviews. To determine whether the sample will support assessment of the impact of the screening and interview incentive conditions on nonresponse bias, SAMHSA conducted a power analysis for two key analyses planned. The goal of the power analysis was to determine minimum detectable differences (MDD) for selected outcomes from the combinations of the two screening incentive amounts (\$0 and \$5) and the two interview incentive amounts (\$30 and \$50). Given that the experiment will test all combinations of these screening and interview incentive amounts, the power analysis examined both marginal mean differences (main effects applicable when the interaction between the screening and interview incentive amounts are not significant) and conditional mean differences (applicable when the interaction of the screening and interview incentive amounts is significant).

The power analysis was conducted for the following outcomes:

- 1) Weighted SRR;
- 2) Weighted IRR; and
- 3) Selected demographic characteristics of household members from the screener.

Several important assumptions were applied to the power analysis for the primary goals of the study:

- 1) The current sample design involves 12,774 SDUs yielding 8,110 completed screenings and 4,000 completed interviews.
- 2) Random allocation of the screening and interview incentive amounts among the SDUs will be equal, so that:
 - 50% of SDUs will be assigned to no screening incentive and 50% assigned to a \$5 screening incentive
 - 50% of SDUs will be assigned to a \$30 interview incentive and 50% assigned to a \$50 interview incentive
- 3) The statistical power assumed for detecting differences in outcomes was 0.80.
- 4) In addition to the standard significance level $\alpha = 0.05$, additional alpha levels were included in the power analysis to observe how MDDs changed across different significance levels. For both marginal differences (main effects applicable when the interaction is not significant) and conditional differences (applicable when the interaction is significant), the power analysis was conducted with significance level $\alpha = 0.05$ and $\alpha = 0.10$. For conditional differences applicable when the interaction is significant, two additional alpha levels were applied: $\alpha = 0.01$ and $\alpha = 0.02$.
- 5) Appropriate design effects were included in all power analysis calculations. The appropriate NSDUH design effects for each outcome were divided by the unequal weighting effect for states, based on the assumption of a national sample of SSRs and NSDUH age group allocation (i.e., states will be sampled proportional to size). By using the NSDUH design effect adjusted for the disproportionate sampling of states,

the impact of clustering is included in the design effect for the SRRs, IRRs, and screener demographic items.

Analysis 1: Weighted SRRs

For *Analysis 1*, the assumed sample size was based on the expected number of eligible SDUs determined in the sample, which was projected to be 10,798.

The null (H_0) and alternative (H_a) hypotheses specified for each marginal difference (main effects applicable when the interaction between the screening and interview incentive amounts are not significant) were:

- 1) Screening incentive and SRR
 H_0 : $SRR(\$5) - SRR(\$0) = 0$
 H_a : $SRR(\$5) - SRR(\$0) > 0$
- 2) Screening incentive and IRR
 H_0 : $IRR(\$5) - IRR(\$0) = 0$
 H_a : $IRR(\$5) - IRR(\$0) > 0$

The null (H_0) and alternative (H_a) hypotheses specified for each conditional difference (applicable when the interaction the screening and interview incentive amounts is significant), were:

- 1) Screening incentive and SRR, conditional on \$30 interview incentive
 H_0 : $SRR(\$5; \$30) - SRR(\$0; \$30) = 0$
 H_a : $SRR(\$5; \$30) - SRR(\$0; \$30) > 0$
- 2) Screening incentive and SRR, conditional on \$50 interview incentive
 H_0 : $SRR(\$5; \$50) - SRR(\$0; \$50) = 0$
 H_a : $SRR(\$5; \$50) - SRR(\$0; \$50) > 0$

Table 2 shows the MDDs for SRRs for marginal mean differences. The table indicates the alpha level assumed for each MDD. For the marginal mean differences, the expected sample size for comparing SRRs across the four incentive conditions will allow for detecting a difference of 4.7% as statistically significant when alpha is 0.05 and a difference of 4.1% as statistically significant when alpha is 0.10.

Table 2. Minimum Detectable Differences for Screening Response Rates: Marginal Mean Differences

<i>Alpha=0.05</i>	<i>Alpha=0.10</i>
0.047	0.041

* Minimal detectable differences in the table represent actual changes in rates. Percentages are presented in the paragraph above.

Table 3 shows the MDDs for SRRs for conditional mean differences. The table indicates the alpha level assumed for each MDD. For the conditional mean differences, the expected sample size for comparing SRRs across the four incentive conditions will allow for detecting differences ranging from 5.6% to 8.3% as statistically significant, across the four alpha levels.

Table 3. Minimum Detectable Differences for Screening Response Rates: Conditional Mean Differences (sample size = n/2 if interaction is statistically significant)

<i>Alpha=0.01</i>	<i>Alpha=0.02</i>	<i>Alpha=0.05</i>	<i>Alpha=0.10</i>
0.083	0.076	0.066	0.056

* Minimal detectable differences in the table represent actual changes in rates. Percentages are presented in the paragraph above.

Analysis 2: Weighted IRRs

For IRRs, sample sizes for the three sampling scenarios were based on the expected number of persons aged 12 or older selected for an interview, among all completed screenings for each sample size scenario, which was projected to be 6,206.

The null (H_0) and alternative (H_a) hypotheses specified for each marginal difference (main effects applicable when the interaction between the screening and interview incentive amounts are not significant) were:

- 1) Interview incentive and SRR
 - H_0 : SRR (\$30) – SRR (\$50) = 0
 - H_a : SRR (\$50) – SRR (\$30) > 0

- 2) Interview incentive and IRR
 - H_0 : IRR (\$30) – IRR (\$50) = 0
 - H_a : IRR (\$50) – IRR (\$30) > 0

The null (H_0) and alternative (H_a) hypotheses specified for each conditional difference (applicable when the interaction the screening and interview incentive amounts is significant), were:

- 1) Interview incentive and IRR, conditional on \$0 screening incentive

$$H_0: \text{IRR}(\$50; \$0) - \text{IRR}(\$30; \$0) = 0$$

$$H_a: \text{IRR}(\$50; \$0) - \text{IRR}(\$30; \$0) > 0$$

- 2) Interview incentive and IRR, conditional on \$5 screening incentive

$$H_0: \text{IRR}(\$50; \$5) - \text{IRR}(\$30; \$5) = 0$$

$$H_a: \text{IRR}(\$50; \$5) - \text{IRR}(\$30; \$5) > 0$$

Table 4 shows the MDDs for IRRs for marginal mean differences. The table indicates the alpha level assumed for each MDD. For conditional mean differences, the expected sample size for comparing IRRs across the four incentive conditions will allow for detecting a difference of 5.2% as statistically significant when alpha is 0.05 and a difference of 4.4% as statistically significant when alpha is 0.10.

Table 4. Minimum Detectable Differences for Interview Response Rates: Marginal Mean Differences

<i>Alpha=0.05</i>	<i>Alpha=0.10</i>
0.052	0.044

Table 5 shows the MDDs for IRRs for conditional mean differences. The table indicates the alpha level assumed for each MDD. For conditional mean differences, the expected sample sizes for comparing IRRs across the four incentive conditions will allow for detecting differences ranging from .2% to 9.2% as statistically significant, across the four alpha levels.

Table 5. Minimum Detectable Differences for Interview Response Rates: Conditional Mean Differences (sample size = n/2 if interaction is statistically significant)

<i>Alpha=0.01</i>	<i>Alpha=0.02</i>	<i>Alpha=0.05</i>	<i>Alpha=0.10</i>
0.092	0.084	0.072	0.062

* Minimal detectable differences in the table represent actual changes in rates. Percentages are presented in the paragraph above.

Analysis 3: Selected Demographic Characteristics of Household Members from the Screener

For the screener demographics data, the sample size was defined as 100% of the projected number of persons aged 12 or older for whom screener data is expected to be collected. This sample size did not include an adjustment for item missingness because the NSDUH screener data for each person residing in a screened household is typically complete. The expected sample sizes for the three demographic items – gender, age, and race/ethnicity – was projected to be 17,390.

The null (H_0) and alternative (H_a) hypotheses specified for each marginal difference (main effects applicable when the interaction between the screening and interview incentive amounts are not significant) were:

- 1) Screening incentive and demographic screener items

H_0 : estimated % (\$5) = estimated % (\$0)

H_a : estimated % (\$5) \neq estimated % (\$0)

- 2) Interview incentive and demographic screener items

H_0 : estimated % (\$30) = estimated % (\$50)

H_a : estimated % (\$30) \neq estimated % (\$50)

The null (H_0) and alternative (H_a) hypotheses specified for each conditional difference, (applicable when the interaction the screening and interview incentive amounts is significant), were:

- 1) Screening incentive and demographic screener items, conditional on \$30 interview incentive

H_0 : estimated % (\$5; \$30) = estimated % (\$0; \$30)

H_a : estimated % (\$5; \$30) \neq estimated % (\$0; \$30)

- 2) Screening incentive and demographic screener items, conditional on \$50 interview incentive

H_0 : estimated % (\$5; \$50) = estimated % (\$0; \$50)

H_a : estimated % (\$5; \$50) \neq estimated % (\$0; \$50)

- 3) Interview incentive and demographic screener items, conditional on \$0 screening incentive

H_0 : estimated % (\$50; \$0) = estimated % (\$30; \$0)

H_a : estimated % (\$50; \$0) \neq estimated % (\$30; \$0)

- 4) Interview incentive and demographic screener items, conditional on \$5 screening incentive

H_0 : estimated % (\$50; \$5) = estimated % (\$30; \$5)

H_a : estimated % (\$50; \$5) \neq estimated % (\$30; \$5)

Table 6 shows the MDDs for the selected demographic screener data for marginal mean differences. The demographic data includes age group, gender (represented as proportion who are men³), and racial and ethnic groups. The table indicates the alpha level assumed for each MDD.

Age groups: For marginal mean differences in Table 6, the expected sample size for comparing age groups across the four incentive conditions will allow for detecting differences ranging from 1.6% to 4.0% as statistically significant, across the six age categories and the two alpha levels.

Gender: For marginal mean differences in Table 6, the expected sample size for comparing gender across the four incentive conditions will allow for detecting a difference of 3.3% as statistically significant when alpha is 0.05 and a difference of 2.9% as statistically significant when alpha is 0.10.

Race/ethnicity groups: For marginal mean differences in Table 6, the expected sample size for comparing race/ethnicity groups across the four incentive conditions will allow for detecting differences ranging from 2.5% to 5.0% as statistically significant, across the four categories and the two alpha levels.

Table 6. Minimum Detectable Differences for Age, Gender, and Race/Ethnicity: Marginal Mean Differences

	<i>Alpha=0.05</i>	<i>Alpha=0.10</i>
<i>Age Group</i>		
Age 12-20	0.021	0.019
Age 21-25	0.018	0.016
Age 26-34	0.027	0.023
Age 35-49	0.029	0.037
Age 50-64	0.037	0.033
Age 65+	0.040	0.035
<i>Gender</i>		
Male	0.033	0.029
<i>Race/Ethnicity</i>		
Non-Hispanic White	-0.050	-0.045
Non-Hispanic Black	0.036	0.031
Non-Hispanic Other	0.028	0.025
Hispanic	0.041	0.036

* Minimal detectable differences in the table represent actual changes in rates. Percentages are presented in the paragraph above.

Table 7 shows the MDDs for the selected demographic screener data for conditional mean differences. The demographic data includes age group, gender (represented as proportion who

³ Using male or female in the power calculations would produce identical results, because this variable is being treated as binary.

are men⁴), and racial and ethnic groups. The table indicates the alpha level assumed for each MDD.

Age groups: For conditional mean differences in Table 7, the expected sample size for comparing age groups across the four incentive conditions will allow for detecting differences ranging from 2.2% to 7.1% as statistically significant, across the six age categories and the four alpha levels.

Gender: For conditional mean differences in Table 7, the expected sample size for comparing gender across the four incentive conditions will allow for detecting differences ranging from 4.1% to 5.6% as statistically significant, across the four alpha levels.

Race/ethnicity groups: For conditional mean differences in Table 7, the expected sample size for comparing race/ethnicity groups across the four incentive conditions will allow for detecting differences ranging from 3.6% to 8.8% as statistically significant, across the four categories and the four alpha levels.

Table 7. Minimum Detectable Differences for Age, Gender, and Race/Ethnicity: Conditional Mean Differences (sample size = n/2 if interaction is statistically significant)

	<i>Alpha=0.01</i>	<i>Alpha=0.02</i>	<i>Alpha=0.05</i>	<i>Alpha=0.10</i>
<i>Age Group</i>				
Age 12-20	0.038	0.035	0.030	0.027
Age 21-25	0.031	0.029	0.025	0.022
Age 26-34	0.047	0.043	0.038	0.033
Age 35-49	0.051	0.047	0.041	0.037
Age 50-64	0.065	0.060	0.052	0.046
Age 65+	0.071	0.065	0.057	0.050
<i>Gender</i>				
Male	0.056	0.052	0.046	0.041
<i>Race/Ethnicity</i>				
Non-Hispanic White	-0.088	-0.081	-0.071	-0.063
Non-Hispanic Black	0.064	0.059	0.051	0.045
Non-Hispanic Other	0.051	0.047	0.041	0.036
Hispanic	0.072	0.066	0.058	0.051

* Minimal detectable differences in the table represent actual changes in rates. Percentages are presented in the paragraph above.

⁴ Using male or female in the power calculations would produce identical results, because this variable is being treated as binary.

Summary of Power Analysis Results:

Focusing on the MDDs for marginal mean differences, an increase of approximately 5% for either the screening or interview response rate will be considered meaningful. The sample and experimental design will be able to detect differences in SRRs between incentive conditions of at least 4.7% and differences in IRRs of at least 5.2% with 80% power and assuming alpha is 0.05. As a result, observed differences between incentive conditions that are at, or above, these differences in rates will be interpreted as statistically significant and meaningfully different. For the demographic composition of screened SDUs, households offered the \$5 screening incentive compared to those not offered the screening incentive, if one or more of the demographic characteristics (1) differs significantly between the no incentive and \$5 incentive condition and (2) the estimate from the \$5 incentive condition is closer to American Community Survey (ACS) estimates, these differences will also be interpreted as statistically significant and meaningfully different. For age groups, marginal mean differences ranging from 1.8% to 4.0% would be detectable as statistically significant. For gender, a marginal mean difference of 3.3% would be statistically significant. For race/ethnicity, marginal mean differences ranging from 2.8% to 5.0% would be statistically significant.

The sample size implemented, the alpha levels used in the analysis, and the significance (or lack thereof) of the interaction between the screening and interview incentives will determine the actual MDDs when comparing outcomes across experimental conditions. As shown in the tables providing the power calculations for conditional mean differences, if the interaction between the screening and interview incentives is significant, this effectively reduces the sample size by half. In this case, the MDDs that can be interpreted as statistically significant and meaningfully different will be higher.

References

Groves, R. M., Couper, M. P., Presser, S., Singer, E., Tourangeau, R., Acosta, G. P., & Nelson, L. (2006). Experiments in producing nonresponse bias. *Public Opinion Quarterly*, 70(5), 720-736.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly*, 68(1), 2-31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency theory of survey participation - Description and an illustration. *Public Opinion Quarterly*, 64(3), 299-308.

Singer, E., & Ye, C. (2013). The use and effects of incentives in surveys. *The Annals of the American Academy of Political and Social Science*, 645, 112–141.

Williams, D., & Brick, M. (2018). Trends in U.S. face-to-face household survey nonresponse and level of effort. *Journal of Survey Statistics and Methodology*, 6, 186-211.