Attachment FT-1

The 2020 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 assess whether the impact of the combination of the levels of the different screening and interview incentive conditions on outcomes (SRR, IRR and demographic compositions of households screened) are significant. If this interaction is not statistically significant, then the incentive experiment will be used to assess the marginal effect of:

1) a screening incentive on screening response rates (SRR);
2) an increased interview incentive on interview response rates (IRR);
3) a screening incentive on nonresponse bias by examining the demographic composition of households screened.[1]

The screening and interview incentive experiment will also be used to assess:

1) an increased interview incentive on SRR;[1]
2) an increased 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) a screening incentive on the number of contact attempts per completed screening; and
4) a screening incentive or increased interview incentive on the number of contact attempts per completed interview.

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

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

Similar to the respondent universe for the annual NSDUH main study, the respondent universe for the FT is the civilian, noninstitutionalized U.S. population aged 12 or older. To control costs, individuals residing in Alaska and Hawaii will be excluded from the FT. Unlike the main study, only respondents who can complete the screening and interview in English will be included in the FT. Approximately 356 segments and 12,774 SDUs will be needed to yield approximately 8,110 completed screening interviews and approximately 4,000 completed interviews. State sampling regions (SSRs), defined as contiguous groups of census tracts in the main NSDUH study, will be used as primary sampling units (PSUs) in the FT. To achieve representation of the age-eligible, English-speaking population in the contiguous United States, a probability proportional to size (PPS) sample of 89 (of 726) of the NSDUH SSRs will be selected. Four segments will be selected in each SSR. The age allocation for interviews will be the same as the current NSDUH main study: 25 percent aged 12 to 17, 25 percent aged 18 to 25, and 50 percent aged 26 or older.

Some important components of the experimental design to note are:

- 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 FT 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 FT field period.

This experimental design assigns 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. This will also 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. Power Analysis

To determine whether the 2020 Redesign FT sample will support assessment of the impact of the screening and interview incentive conditions on nonresponse bias, SAMHSA conducted a power analysis for 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 conditional mean differences (applicable when the interaction of the screening and interview incentive amounts is significant) and marginal mean differences (main effects applicable when the interaction between the screening and interview incentive amounts are not significant).

The power analysis was conducted for the following outcomes:

1) Weighted FT SRR;
2) Weighted FT IRR; and
3) Selected demographic characteristics of household members from the FT 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 α = 0.05, additional alpha levels were included in the power analysis to observe how MDDs changed across different significance levels. For marginal differences (main effects applicable when the interaction is not significant), the power analysis was conducted with significance level α = 0.05. For conditional differences (applicable when the interaction is

4

significant), the power analysis was conducted with significance levels $\alpha = 0.05$ and $\alpha = 0.10$.

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 FT sample.

The null (H_o) 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_o: SRR ($5) – SRR ($0) = 0

H_a: SRR ($5) – SRR ($0) > 0

2) Screening incentive and IRR

H_o: IRR ($5) – IRR ($0) = 0

H_a: IRR ($5) – IRR ($0) > 0

The null (H_o) 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_o: 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_o: SRR ($5; $50) – SRR ($0; $50) = 0

H_a: SRR ($5; $50) – SRR ($0; $50) > 0

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

Table 2 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 6.6% as statistically significant, across the two alpha levels.

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

Alpha=0.05	Alpha=0.10
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, the power analysis was based on the expected number of persons aged 12 or older selected for an interview, among all completed screenings.

The null (H_o) 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_o: SRR ($30) – SRR ($50) = 0

   H_a: SRR ($50) – SRR ($30) > 0

2) Interview incentive and IRR

   H_o: IRR ($30) – IRR ($50) = 0

   H_a: IRR ($50) – IRR ($30) > 0

The null (H_o) 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_o: IRR ($50; $0) – IRR ($30; $0) = 0

   H_a: IRR ($50; $0) – IRR ($30; $0) > 0

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

   H_o: IRR ($50; $5) – IRR ($30; $5) = 0

   H_a: IRR ($50; $5) – IRR ($30; $5) > 0

For marginal 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.

Table 3 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 6.2% to 7.2% as statistically significant, across the two alpha levels.

Table 3. Minimum Detectable Differences for Interview Response Rates: Conditional Mean Differences

Alpha=0.05	Alpha=0.10
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 null (H_o) 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_o: estimated % ($5) = estimated % ($0)

   H_a: estimated % ($5) ≠ estimated % ($0)

2) Interview incentive and demographic screener items

H$_o$: estimated % ($30) = estimated % ($50)

H$_a$: estimated % ($30) ≠ estimated % ($50)

The null (H$_o$) 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$_o$: estimated % ($5; $30) = estimated % ($0; $30)

    H$_a$: estimated % ($5; $30) ≠ estimated % ($0; $30)

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

    H$_o$: estimated % ($5; $50) = estimated % ($0; $50)

    H$_a$: estimated % ($5; $50) ≠ estimated % ($0; $50)

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

    H$_o$: estimated % ($50; $0) = estimated % ($30; $0)

    H$_a$: estimated % ($50; $0) ≠ estimated % ($30; $0)

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

    H$_o$: estimated % ($50; $5) = estimated % ($30; $5)

    H$_a$: estimated % ($50; $5) ≠ estimated % ($30; $5)

Table 4 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 men[2]), and racial and ethnic groups. The table indicates the alpha level assumed for each MDD.

Age groups: For marginal mean differences in Table 4, the expected sample size for comparing age groups across the four incentive conditions will allow for detecting differences ranging from 1.8% to 4.0% as statistically significant, across the six age categories.

Gender: For marginal mean differences in Table 4, the expected sample size for comparing gender across the four incentive conditions will allow for detecting a difference of 3.3% as statistically significant.

Race/ethnicity groups: For marginal mean differences in Table 4, the expected sample size for comparing race/ethnicity groups across the four incentive conditions will allow for detecting differences ranging from 2.8% to 5.0% as statistically significant, across the four categories.

Table 4. Minimum Detectable Differences for Age, Gender, and Race/Ethnicity: Marginal Mean Differences

Demographic Category	Alpha=0.05
Age Group	
Age 12-20	0.021
Age 21-25	0.018
Age 26-34	0.027
Age 35-49	0.029
Age 50-64	0.037
Age 65+	0.040
Gender	
Male	0.033
Race/Ethnicity	
Non-Hispanic White	-0.050
Non-Hispanic Black	0.036
Non-Hispanic Other	0.028
Hispanic	0.041

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

Table 5 shows the MDDs for the selected demographic screener data for conditional mean differences. The demographic data includes age group, gender (represented as proportion who are men[3]), and racial and ethnic groups. The table indicates the alpha level assumed for each MDD.

---

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

Age groups: For conditional mean differences in Table 5, the expected sample size for comparing age groups across the four incentive conditions will allow for detecting differences ranging from 2.2% to 5.7% as statistically significant, across the six age categories and the two alpha levels.

Gender: For conditional mean differences in Table 5, the expected sample size for comparing gender across the four incentive conditions will allow for detecting differences ranging from 4.1% to 4.6% as statistically significant, across the two alpha levels.

Race/ethnicity groups: For conditional mean differences in Table 5, the expected sample size for comparing race/ethnicity groups across the four incentive conditions will allow for detecting differences ranging from 3.6% to 7.1% as statistically significant, across the four categories and the two alpha levels.

Table 5 Minimum Detectable Differences for Age, Gender, and Race/Ethnicity: Conditional Mean Differences

Demographic Category	Alpha=0.05	Alpha=0.10
Age Group		
Age 12-20	0.030	0.027
Age 21-25	0.025	0.022
Age 26-34	0.038	0.033
Age 35-49	0.041	0.037
Age 50-64	0.052	0.046
Age 65+	0.057	0.050
Gender		
Male	0.046	0.041
Race/Ethnicity		
Non-Hispanic White	-0.071	-0.063
Non-Hispanic Black	0.051	0.045
Non-Hispanic Other	0.041	0.036
Hispanic	0.058	0.051

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

***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 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 FT 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. 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.

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.