

Power Analysis Details for NSDUH Incentives Experiment

Key Assumptions Applied to the SRR and IRR Power Calculations

- 1) The power analysis assumed a difference of at least 5% in SRRs or IRRs between test and control incentive conditions would be statistically and practically meaningful for inferring a positive improvement in response rates from the increased incentive amounts.
- 2) The power analysis was conducted using two alpha (significance) levels for marginal mean differences:
 - $\alpha = 0.05$
 - $\alpha = 0.10$

Note that the SDUs presented in Table 2.3 and the selected interview respondents presented in Table 2.4 are based on the calculations assuming $\alpha = 0.05$. Focusing on $\alpha = 0.05$ was intended to ensure a sufficient number of Quarter 4 2022 SDUs and selected interview respondents will be selected for each of the three test incentive conditions. This approach also conforms with the conventional significance level used for statistical inference.

Table 2.3 Sample Dwelling Units for the Four Experimental Conditions for the Incentives Experiment

	\$0 Screener Incentive	\$5 Screener Incentive	Interview Incentive Marginal Totals
\$30 Interview Incentive	remainder	6,250	>12,500
\$50 Interview Incentive	6,250	6,250	12,500
Screener Incentive Marginal Totals	>12,500	12,500	

Table 2.4 Selected Interview Respondents for the Four Experimental Conditions for the Incentives Experiment

	\$30 Interview Incentive	\$50 Interview Incentive	Screening Incentive Marginal Totals
No Screening Incentive	Remainder of sample	2,100	>4,200
\$5 Screening Incentive	2,100	2,100	4,200
Interview Incentive Marginal Totals	>4,200	4,200	

- 3) One-sided tests were assumed, which are directional. The effect in the desired direction was used. That is, the null hypothesis of no effect from increased incentives is rejected if the response rates with higher incentives are at least five percentage points larger than response rates under current incentive levels.
- 4) The statistical power assumed was 0.80.
- 5) The degrees of freedom were assumed to be 60. The degrees of freedom correspond to the number of variance strata, which should be larger than 60. However, anything over 60 has very little impact on the results, so 60 is often used for simplicity.
- 6) The design effects (DEFFs) applied to the power calculations accounted for clustering of the treatment cases and unequal weighting effect at the DU level (i.e., because DUs are selected within segments) under the current state-based design. The DEFFs for SRR sample size calculations were 4.69 with $\alpha = 0.05$ and 4.18 with $\alpha = 0.10$. The DEFFs for IRR sample size calculations were 3.04 with $\alpha = 0.05$ and 3.01 with $\alpha = 0.10$. Note that the larger alpha requires a smaller sample size in each treatment group. In turn, this means less clustering of treatment cases and a smaller design effect.
- 7) The DU eligibility was assumed to be 85%, the control screener rate was assumed to be 27%, and the persons selected per DU was assumed to be 68.6%.
- 8) For calculations, the treatment and control groups for each test were expected to have the same number of SDUs. Because the experiment involves four treatment conditions in a 2x2 experiment (Table 2.2), the experiment requiring the larger sample sizes would determine the sample sizes for both tests. The incentives experiment is embedded into production data collection, so the number of SDUs in Q4 of 2022 will far exceed the number needed for both experiments. All excess SDUs will be assigned to the current control condition.

As shown in the second table below, the power analysis calculations indicated a total of 11,436 SDUs ($22,871 \div 2$) will need to be assigned to each interview incentive condition. This exceeds the number needed for the screener experiment, so the same number is then assigned to the screener incentive conditions, too. That is, each marginal total is 11,436. Taking into consideration the potential for one or more key assumptions to prove incorrect and other unforeseen circumstances, RTI recommended the higher number of 12,500 SDUs to ensure sufficient sample sizes for the incentives experiment. This number of SDUs is presented in the marginal total SDUs presented in Table 2.3. The marginal total SDUs were also used to derive the selected interview respondents needed in Table 2.4.

Table A.1: Assumptions and Calculations for the SRR Sample Sizes

Alpha:	0.05	0.10
Probabilities:		
Null Hypothesis - No Treatment Effect	0.270	0.270
Alternative hypothesis for treatment group	0.320	0.320
Difference (minimum detectable effect)	0.050	0.050
Combined (for null hypothesis variance)	0.295	0.295
Sampling Allocation:		
Null hypothesis proportion	0.500	0.500
Alternative hypothesis proportion	0.500	0.500
Variance Calculation:		
Null hypothesis variance	0.8316	0.8312
Alternative hypothesis variance	0.8290	0.8287
Test Information:		
Alpha	5%	10%
# of sides in test	1	1
Power	80%	80%
Degrees of freedom	60	60
t-alpha	1.671	1.296
t-beta	0.848	0.848
Design Effect (DEFF):		
DU DEFF	4.69	4.18
Assumed DU Eligibility Rate (from 2019):	0.85	0.85
Sample Size (<i>n</i>) Calculations:		
Effective <i>n</i> required (eligible DUs for attempted screening)	2,074	1,526
Actual <i>n</i> required (accounts for DEFF)	9,729	6,380
Initial <i>n</i> required (accounts for eligibility, conservatively)	11,384	7,466

Notes:

1. Formula for sample size calculation for two proportions (using normal approximation) based on Fleiss (1981), but modified to use the *t*-distribution rather than the standard normal *z*-distribution.
2. SRR and IRR null hypothesis values and associated design effects were obtained from Q4 2021 NSDUH.
3. Eligibility rate was obtained from the 2019 NSDUH, to be more conservative than current mixed sample assumptions.

Table A.2: Assumptions and Calculations for the IRR Sample Sizes

Alpha:	0.05	0.10
Probabilities:		
Null Hypothesis - No Treatment Effect	0.456	0.456
Alternative hypothesis for treatment group	0.506	0.506
Difference (minimum detectable effect)	0.050	0.050
Combined (for null hypothesis variance)	0.481	0.481
Sampling Allocation:		
Null hypothesis proportion	0.500	0.500
Alternative hypothesis proportion	0.500	0.500
Variance Calculation:		
Null hypothesis variance	0.9986	0.9986
Alternative hypothesis variance	0.9961	0.9961
Test Information:		
Alpha	5%	10%
# of sides in test	1	1
Power	80%	80%
Degrees of freedom	60	60
t-alpha	1.671	1.296
t-beta	0.848	0.848
Design Effect (DEFF):		
Person DEFF	3.04	3.01
Person Eligibility Rate:	1.0	1.0
Sample Size (n) Calculations:		
Effective <i>n</i> required (total selected persons)	2,531	1,834
Actual <i>n</i> required (accounts for DEFF)	7,684	5,519
Initial <i>n</i> required (accounts for person eligibility)	7,684	5,519
Expected persons selected per screened DU	0.686	0.686
Screened DUs required	5,269	3,784
Assumed Screener Rate	0.270	0.270
Eligible DUs required	19,546	14,037
DU Eligibility Rate	0.85	0.85
Initial sample DUs required	22,871	16,425

Notes:

1. Formula for sample size calculation for two proportions (using normal approximation) based on Fleiss (1981), but modified to use the t -distribution rather than the standard normal z -distribution.
2. SRR and IRR null hypothesis values and associated design effects were obtained from Q4 2021 NSDUH.
3. Eligibility rate was obtained from the 2019 NSDUH, to be more conservative than current mixed sample assumptions.