

Supporting Statement
Internal Revenue Service
Statistics of Income Division
Appendix B
Comprehensive Taxpayer Attitude Survey
August 2023 through August 2026
Triennial Approval Request

B. Collections of Information Employing Statistical Methods.

1. Universe and Respondent Selection

Individuals from the general public over the age of 18 that match the demographic characteristics of the U.S. population, and whom we assume are taxpayers, will be reached via technology-enhanced address-based sampling (ABS) and dynamic online targeting and will be invited to participate in a survey using an automated interactive voice response (IVR) phone survey or an online survey.

2. Procedures for Collecting Information

CTAS continues the previous multi-mode data collection methodology, comprised of randomly sampled telephone and online surveys, to ensure a representative sample of the general public. However, we will replace Random Digit Dialed (RDD), interviewer-led phone surveys and online panel surveys with automated IVR surveys and online surveys recruited by mail, through ABS, and online advertisements at sites hard-to-reach populations use. The vendor will collect a total of 1,298 completed surveys, split between the telephone and online data collection methodologies according to respondents' choices, for a margin of error (MOE) of +/- 3% at the 95% confidence level. The survey questionnaire has been streamlined and refined based on Office of Management and Budget (OMB)-approved focus groups that identified questions more relevant to taxpayer's concerns and challenges.

Potential respondent universe

Adult population (18 years or older) in the United States

Numerical Estimate of the respondent universe

258,672,000 based on 2021 population projections from US Census Bureau for 2017 to 2060.¹

¹ <https://www2.census.gov/programs-surveys/popproj/tables/2017/2017-summary-tables/np2017-t2.xlsx>

Number of Entities

Data Collection Method	Potential Respondent Universe (adults 18+ in US)	Requested for participation	Participants in the survey (survey sample size)
Online Recruitment: Online Response	258,672,000	32,450	649
Mail Recruitment: Phone Response	258,672,000	32,450	130
Mail Recruitment: Online Response	258,672,000	32,450	519
Overall Total	258,672,000	32,450	1,298

Methodology

Respondent Selection Method: Technology-enabled random address-based (ABS) sampling uses a comprehensive sampling frame with technological strategies to maximize response rates, particularly among groups underrepresented by other methods. Our vendor team applies innovations to the scientific method of ABS. To ensure a representative sample, hard-to-reach respondents are identified, over-sampled, and sent customized invitations. For example, they use an algorithm to combine information on surnames and neighborhood to predict the most likely language spoken in the home. Households that are likely to be Spanish speaking are over-sampled and sent bilingual invitations in English as well as Spanish. Targeted populations for the 2023 survey include Spanish Limited English Proficiency (LEP), Asian-Language LEP, Homeless, Rural/Low-Income, Stateside Military, and Veterans. Homeless individuals will be contacted through homeless shelters.

Tech-enabled ABS solves challenges of the previous CTAS methods, but it also allows IRS to reduce costs, while maintaining data quality, when the ABS sample is combined with dynamic online targeting, which reaches potential respondents at online sites they use. Respondents are recruited through online advertisements in English and Spanish across the web (e.g., social media, search engines, and news websites). This process continually recruits new participants for each survey and does not rely on any third-party online panels. The dynamic online targeting approach can be optimized to recruit hard-to-reach populations, such as dedicated targeting of Americans whose browsers are set to Spanish. Further, the targeting technology adjusts as the sample grows, ensuring that the final sample is representative of the desired population. The method is “dynamic” because respondents are recruited continuously, and it is “targeted” because the technology

adjusts to ensure the sample is representative of the desired population. Our vendor team's use of tech-enabled ABS and dynamic online targeting improves representation, eliminates concerns of panel fatigue, and significantly reduces costs.

Representative Sample: In order to achieve survey data representative of the U.S. adult online population, our vendor team uses several unique features to ensure a scientific sample that is representative of the national population, including those with limited English proficiency, non-internet users, low-income individuals, and the homeless. Its sampling frame is based on all available household addresses identified through the United States Postal Service. The primary sources for the household data come from the DSF2[®], which identifies whether a ZIP + 4[®] coded address is currently represented in the USPS[®] delivery file as a known address record, and from the NCOALink[®], which represents the change-of-address (COA) records consisting of the names and addresses of individuals, families and businesses who have filed a change-of-address with the USPS. These sources, along with their other address-based information, offer a comprehensive sampling frame of all U.S. households.

From this complete list of U.S. addresses, they take a random sample of all households to generate a secondary sampling frame that represents all U.S. households. With this secondary sampling frame, they combine local area demographic information (e.g., the percentage of the local population of various demographic characteristics) as well as the name of the household to estimate a probability that the household is Spanish speaking, low-income, or a member of another hard-to reach group. They then oversample these key hard-to-reach groups and send additional mailers, in some cases. This approach ensures that the sample reflects the entire U.S. population and ensures representation of hard-to-reach groups such as those with limited English proficiency or those from lower socioeconomic backgrounds.

The tech-enabled ABS approach begins with a letter invitation to all individuals in the sample frame. They then send a follow-up mailer approximately 5 days after the initial invitation. Importantly, outreach and timing are customized to recruit hard-to-reach populations. The vendor team's models use surname and address to predict language spoken in the household, income, and other characteristics that might be correlated with response rate. For example, individuals identified as likely Spanish speakers will receive a bilingual survey invitation and other hard-to-reach groups like those who are homeless, might be given an additional incentive. The questionnaire could be used to further classify Spanish speakers into LEP-Spanish, mirroring the U.S. Census approach asking: Does the person speak a language other than English at home? Which language? How well does the person speak English (very well, well, not well, not at all)?

Respondents have three self-directed options to complete the survey. Each mailer includes a quick response (QR) code, which can be scanned with a smartphone, a URL which can be typed into a computer, and a 1-800 number for those who want to call to take the survey. The 1-800 option ensures that individuals without access to the internet

can take the survey. Further, because these individuals have the option to call, they can take the survey whenever it is convenient. These individuals have the option to take the survey via IVR or from a live interviewer, whichever they prefer.

Using the sampling methods proposed, the vendor team's previous ABS response rate has been above 3%, which is higher than recent CTAS random-digit dial response rates.

Weighting Strategy

The online and phone data will be combined into an aggregate data set and then weighted to the U.S. population. The vendor will weight the final sample by standard demographic variables as well as by the proportion of the population with access to the internet to ensure the full sample is completely representative of the U.S. population. This includes a raking technique to generate initial base weights to match Census benchmarks of key demographic variables. After calculating these base weights, they will apply a multilevel regression with synthetic poststratification procedure to a key survey item. The results from this model allow the vendor to adjust the base weights to create final weights that incorporate information about a) the interactions between weighting variables, and b) the nested nature (e.g., within geographic units) of the survey data. This additional weighting step provides increased accuracy beyond standard raking approaches.

Analytical Planning

The vendor will review an interim survey dataset in the latter stage of fielding. The dataset will be cleaned and basic frequencies for each question will be generated. A review of this initial analysis in addition to the vendor's experience analyzing previous years' data will enable the vendor's analysis and reporting plan for the final dataset(s). The vendor will share this plan with RAAS, and they will meet to discuss the proposed plan and whether it sufficiently meets RAAS's needs.

3. Methods to Maximize Response

Offering different survey modes that fit the preferences of different segments helps to maximize response. Respondents recruited via address-based sampling have the option to respond via web or by IVR phone surveys. Allowing respondents to participate in their preferred language, at their preferred time, and in their preferred mode increases response rates and response validity. Additionally, incentives will be offered to those who complete the survey in the form of \$10 gift cards, which will be given to the homeless even for completing a partial survey.

Fielding is planned for 4 weeks for the online survey, similar to past years, which should allow adequate time to participate. The vendor will concurrently offer a self-administered IVR option. Those who start the IVR and then experience difficulties can push a key and be connected to a live interviewer. Depending on how the on-demand sampling is progressing and on the quality of sample being obtained through IVR or IVR-to-live

interview, we may add 1-2 more weeks to the fielding timeline, which helps to reach underserved segments and encourages responses.

4. Testing Procedures

IRS employees who are familiar with the product or service under investigation review the questionnaires used in the surveys

The vendor will fully brief field staff responsible for data collection and the fieldwork will be monitored. This is an established survey that has been conducted for many years and which reflects the extensive experience of IRS and the vendor in conducting this and similar surveys. The 2023 questionnaire has been streamlined and refined based on focus group input, for which a separate OMB clearance was approved. When appropriate, questionnaires may be pre-tested on up to nine external customers. If significant changes were made, RAAS would provide the revised questionnaire to OMB for review.

5. Contact for Statistical Aspects and Data Collection

Questions regarding any statistical aspects employed or data procedures used, or for administrative questions regarding the IRS use of this clearance should be directed to:

Timothy S. Castle
Internal Revenue Service
Statistics of Income Division
77 K Street, NE
Washington, DC 20002
(202) 803-9106

Donna Baldwin
Internal Revenue Service
Statistics of Income Division
77 K Street, NE
Washington, DC 20002
(571) 516-1052