

**OMB SUPPORTING STATEMENT
INTERNAL REVENUE SERVICE
IRS TAXPAYER BURDEN SURVEYS
TIRNO-10-Q-00152**

The Supporting Statement for OMB 1545-2212

IRS Taxpayer Burden Surveys

PART B - JUSTIFICATION

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used.

Individual Burden Surveys

The potential respondent universe is composed of wage and investment and self-employed taxpayers living in the United States. These taxpayers file a Form 1040, 1040-A, 1040-EZ, 1040-NR, 1040-NR-EZ, or 1040-X (as well as supporting forms and schedules). The sample frame will be developed using IRS administrative data sources, such as the returns transaction file (RTF). Some populations will be explicitly excluded from the survey population. This includes taxpayers that are minors, deceased taxpayers, and taxpayers that have international addresses, including active duty military serving overseas.

When sub-populations vary considerably, it is advantageous to sample each subpopulation (stratum) independently. Stratification is the process of grouping members of the population into relatively homogeneous sub-groups before sampling. The strata should be:

- Mutually Exclusive. Members must be assigned to only one stratum, and
- Collectively Exhaustive. No members can be excluded.

Then, random or systematic sampling can be applied within each stratum. Stratification often improves the representativeness of the sample by reducing sampling error. It also tends to produce a weighted mean that has less variability than the arithmetic mean of a simple random sample of the population. For these reasons, the proposed sample design for this study is a stratified random sample.

The sampling approach has been designed to ensure that key taxpayer subgroups are adequately represented in the study findings. The stratification includes two main criteria:

- Preparation method. The method by which the taxpayer prepared his or her return.
 - Prepared by a paid professional (paid)
 - Prepared using tax preparation software (soft)
 - Prepared by hand (self)
- Differential burden. Variable reflecting type of activities performed by taxpayers to meet their Federal tax obligations. Taxpayers are assigned burden corresponding to the highest burden item reported on their tax forms.

Differential burden is summarized in the following table.

Strata	Definition
Low	Wage income; Interest income; Unemployment income; Withholding; Earning income tax credit (with no qualifying children) or advanced EIC; Does not meet any of the conditions for higher levels of differential burden
Low-Medium	Capital gain income (includes capital gains distributions and undistributed capital gains); Dividend income; Earned income tax credit (with qualifying children); Estimated tax payments; Retirement income (includes SS benefits, IRA distributions, or pensions and annuities); Any non-refundable credit (includes child and dependent care expenses, education credits, child tax credit, elderly or disabled credit); Household employees; Non-business adjustments; Does not meet any of the conditions for higher levels of differential burden
Medium	Itemized deductions (includes mortgage interest, interest paid to financial institution; charitable contributions, and medical expenses); Foreign income, expense, tax, credit, or payment; Moving expenses; Simple Schedule C or C-EZ; General business credit; Does not meet any of the conditions for higher levels of differential burden
Medium-High	Farm income as reported on Schedule F; Owns rental property as reported on Schedule E, including farm rental and low income housing; Estate or trust income as reported on Schedule E; Employee business expense deductions; Files AMT without AMT preference items; Prior year alternative minimum tax credit; Investment interest expense deduction; Net loss as reported on Schedule C; Depreciation or amortization as reported on Schedule C; Expenses for business use of home as reported on Schedule C; Does not meet any of the conditions for higher levels of differential burden
High	Cost of goods sold as reported on Schedule C; Partnership or S-Corp income as reported on Schedule E; Files AMT with AMT preference items

These variables were chosen for stratification because of their importance to the modeling of taxpayer burden and behavioral activities. The differential burden variable is included to ensure that different tax concepts, tax provisions, and tax characteristics with differential recordkeeping and reporting requirements are included. The tax preparation method variable ensures both a proper balance and an adequate representation of paid preparers, software preparers, and self preparers, allowing us to reflect the role of technology and services in meeting recordkeeping and reporting requirements.

The specifications of the sample design are developed to balance three main issues. The first is that it must be efficient in the way the sample is distributed so that estimates from the sample are reliable (i.e., meet confidence interval range requirements). Specifically, the aim is for the coefficient of variation to be under 2%. The second is to ensure that there are a sufficient number of cases to meet the needs of the modeling tool to identify the determinants of burden within and across strata. The third is that the design should facilitate comparisons between future Individual Taxpayer Burden surveys and the previous surveys.

To make the Individual Taxpayer Burden survey comparable with the previous surveys, we continue to use the same design variable (total monetized burden), the same stratified random sampling approach, and the same stratification variables as in the tax year 2007 survey. In the 2007 survey, the Neyman allocation method was used to determine the sample size for each stratum, subject to the total sample size of 15,000. It aimed to minimize the variance of estimated mean burden; however, it limited the sophistication of the modeling of certain thin populations of interest. For the tax year 2010 survey, we adjusted the Neyman allocation by requiring a minimum number of observations per stratum. The minimum number of observations was defined by applying a common rule of thumb, which states that a sample must include at least 10 or 15 observations per independent variable in a regression model (Stevens, 2002; Bartlett et al., 2001). To be conservative, we chose 15. Given that the expected number of independent variables is 15, the minimum desired number of complete responses for modeling each stratum is 225.

Our objective was to minimize the variance of estimated mean burden constrained on this minimum sample size for modeling, with response rate incorporated. We started with the same total sample size of 15,000 as in the tax year 2007 study, considering this as our base sample. We then calculate the coefficient of variation, given the minimum stratum size of 225. Because the coefficient of variation was too large for the sample size of 15,000, we adjusted the sample size to 20,000, and recalculated the coefficient of variation. The sample size of 20,000 resulted in a coefficient of variation of 1.62%. This coefficient of variation met our requirement. Because we used a new data collection protocol, it also allowed us some additional confidence that we will achieve the desired number and mix of complete responses. See Table B1, **Overall Return Distribution by Strata**, below.

Table B1 – Sample allocation for ITB TY2010 survey

Monetized Burden Strata	Projected Pop Count	Est. Mean	Est. Std. Dev.	Est. Response Rate	Sample Allocation	Expected Number of Respondents
11 paid, low	9,822,075	190.46	241.53	0.2558	880	225
12 paid, low-medium	26,114,402	295.10	370.49	0.3213	1,644	528
13 paid, medium	15,940,360	619.92	980.87	0.3916	2,656	1,040
14 paid, medium-high	15,732,824	946.43	1,157.12	0.3970	3,092	1,228
15 paid, high	10,685,596	1,837.13	2,524.26	0.3894	4,582	1,784
21 self, low	3,503,015	85.97	115.25	0.3594	626	225
22 self, low-medium	2,707,918	157.75	225.08	0.3436	655	225
23 self, medium	1,695,808	499.83	709.51	0.4355	517	225
24 self, medium-high	770,422	715.88	876.97	0.4046	556	225
25 self, high	288,597	923.48	881.83	0.4119	546	225
31 soft, low	10,478,344	116.18	159.24	0.3058	736	225
32 soft, low-medium	15,971,640	185.25	228.28	0.3678	619	228
33 soft, medium	10,942,941	518.45	713.67	0.4620	1,327	613
34 soft, medium-high	6,336,666	769.97	1,015.50	0.4396	1,093	480
35 soft, high	1,639,707	1,278.71	1,615.97	0.4772	472	225
Total	132,630,316	551.90			20,000	7,701
Overall CV						1.62%

A similar approach is expected for the TY2011 ITB survey. Using preliminary results from the TY2010 to guide the sample allocation, we expect to be able to further refine the accuracy and efficiency of the information collection. However, not having the benefits of this analysis at this time, we will conservatively assume identical sample size, response rate and CV for the TY2010 and TY2011 ITB surveys.

The 2011 Post-filing Burden Survey Universe is predominately TY2008, TY2009, and TY2010 taxpayers who had either a Collection or Examination case close in Calendar Year 2011. For this study we propose to survey 8,000 taxpayers stratified into lower and higher return complexity and by their subsequent interaction with IRS Examination, Collection and Appeals processes. Not having appropriate post-filing burden variance data available, the 2011 Post-filing burden study will use the variance of post-filing cycle time to stratify the sample as

a proxy. We propose a follow-up 2012 post-filing burden of similar design, which would use burden variance from the 2011 study to derive a more accurate and efficient design.

Entity Burden Surveys

Business entity taxpayers are defined here as corporations, limited liability companies, or partnerships filing any of the following income tax returns: Forms 1065, 1065-B, 1066, 1120, 1120-F, 1120-FSC, 1120-L, 1120-ND, 1120-PC, 1120-RIC, 1120-REIT, 1120-S, or 1120-SF.

Tax-exempt organizations to be surveyed are taxpayers filing any of the Form 990, 990-EZ, or 990-PF. Organizations filing Form 990-N are excluded from this planned survey data collection because their filing requirements and compliance burden are minimal. The sample frame is all such tax exempt returns from the IRS Returns Inventory and Classification System (RICS). Business entities and tax-exempt organizations will be selected using a modified Neyman allocation. The sample strata will be based on return type, preparation method, and total revenue. Preparation and revenue strata are shown below:

Preparation Method Strata:

1. Self Prepared
2. Paid Prepared (defined as presence of a paid preparer)

Total Revenue Strata:

Self Prepared:

1. Less than \$5,000
2. \$5,001 - \$100,000
3. \$100,001 - \$1,000,000
4. \$1,000,001 or more

Paid Prepared:

1. Equal to zero
2. \$1 - \$5,000
3. \$5,001 – 50,000
4. \$50,001 - \$100,000
5. \$100,001 - \$500,000
6. \$500,001 - \$1,000,000
7. \$1,000,001 - \$5,000,000
8. \$5,000,001 - \$10,000,000
9. \$10,000,001 or more

The specifications of the sample design were developed to balance two main issues. The first and most important is to ensure that there are a sufficient number of expected complete responses to meet the needs of the modeling tool to identify the determinants of burden and their relative impacts. The second issue is that the sample design must be efficient in the way the sample is distributed so that estimates from the sample are reliable (i.e., meet confidence range requirements).

As defined above, the three variables that will be used for stratification are: tax form type, tax preparation method, and total revenue. These variables were chosen for stratification because of their importance to the modeling of taxpayer behavior activities. Tax form type is included to ensure that an adequate number of the different form types are included. The tax preparation method ensures both a proper balance and an adequate representation of paid preparers versus self-preparers. Stratifying on total revenue will ensure that an even distribution of different organization sizes will be included.

For each stratification variable, category breaks were chosen to increase the precision of sample estimates. This is accomplished by choosing breakpoints that segment taxpayers into groups whose burden is homogeneous – or alternatively, breakpoints that create separate estimation strata for groups whose burden is highly variable. Sample size can then be concentrated on taxpayer segments with heterogeneous burden, thus increasing the precision of population estimates.

The overall return distribution for the 2010 Tax-Exempt Burden Survey is shown in Table B2 on the next page.
Table B2 – Overall Tax-Exempt Entity Return Distribution by Strata

Preparation Method Strata:

1. Self Prepared	262,629
2. Paid Prepared	398,147

Total Revenue Strata:

Self Prepared:

1. Less than \$5,000	75,398
2. \$5,001 - \$100,000	143,950
3. \$100,001 - \$1,000,000	33,757
4. \$1,000,001 or more	9,524

Paid Prepared:

1. Equal to zero	40,429
2. \$1 - \$5,000	21,568
3. \$5,001 – 50,000	101,127
4. \$50,001 - \$100,000	120,157
5. \$100,001 - \$500,000	35,563
6. \$500,001 - \$1,000,000	49,065
7. \$1,000,001 - \$5,000,000	11,485
8. \$5,000,001 - \$10,000,000	13,588
9. \$10,000,001 or more	5,165

We plan to sample 12,000 TY2010 Tax Exempt taxpayers drawn from the strata above. For the proposed TY2011 Business Taxpayer Burden survey we expect to sample 24,000 taxpayers in a manner similar to that approved for TY2009 under OMB control #1545-1432. While we expect to improve the efficiency of the design of the TY2011 Business Taxpayer Burden survey based on analysis of the TY2009 data, until such time as we conduct that analysis we will conservatively assume a burden impact corresponding to the assumption of the TY2009 Business Taxpayer Burden information collection request.

2. Describe the procedures for the collection of information.

We have two objectives in the design of this protocol. The first is the efficient collection of the current sample; the second is to inform the design of future studies. The exact form of each of these contacts will vary somewhat, depending upon several factors. First, it will depend upon whether the contractor is able to obtain a telephone number for sampled taxpayers. For those respondents able to be matched with phone numbers, communication will take place via mail with telephone follow-up contact, if necessary. For respondents without phone numbers, all communication will necessarily take place via mail. The anticipated success rate for matching the sample to telephone numbers is about 50%.

The survey protocols for the web-first or survey-first options are shown on the next page.

	<input type="checkbox"/> Web-First Option	<input type="checkbox"/> Survey-First Option
Step 1: Invitation to the survey.	A hardcopy letter will be sent to the targeted respondent inviting the individual taxpayer or individual most familiar with the entity’s tax return to go to the website URL to complete the online survey. The invitation will include information about the survey, assurances that there is no risk associated with participation, and web access information. In addition, respondents will be given directions on how to obtain a paper survey if they do not have access to the web or would prefer a hard copy. This mailing will also include a letter from an IRS official endorsing the survey and emphasizing the importance of the data collection effort.	A paper questionnaire will be mailed to the targeted respondent. The paper-and-pencil mail survey will include a postage-paid return envelope. The survey will include a series of frequently asked questions that will provide information about the survey and assurances that there is no risk associated with participation. Respondents will also be informed of the option to complete a web survey. This mailing will also include a letter from an IRS official endorsing the survey and emphasizing the importance of the data collection effort.
Step 2: Mail survey sent.	A paper questionnaire will be mailed to those households who have not responded to either the initial letter invitation within 15 days of the initial mailing. The paper-and-pencil mail survey will include a postage-paid return envelope.	
Step 3: Thank you/Reminder postcard (Nonrespondents not matched with phone numbers)	Approximately 30 days after the initial mailing, a thank you/reminder postcard will be mailed to all respondents, including those who 1) have completed the survey or 2) who have not completed the survey and for whom no contact phone number is available. The postcard will thank those who have already submitted a completed web survey and ask those who have not to please do so. The postcard will include access information for the web survey.	
Step 3: Phone prompt.	If no completed survey is received either by mail or web within approximately 30 days after the initial mailing, non-respondents that have been matched to a phone number will receive a prompt from a telephone interviewer asking them to complete the survey. Telephone interviewers will be prepared at this stage to administer the interview over the telephone if the respondent wishes.	
Step 4: Thank you/second reminder prompt.	Approximately 60 days after the initial mailing, a thank you/reminder postcard will be mailed to all respondents. The postcard will thank those who have already submitted a completed web survey and ask those who have not to please do so. The postcard will include access information for the web survey.	

In the case of undeliverable mail, there are two possible scenarios: (1) a new address is available and supplied by the post office and (2) no new address is available. For each scenario, a detailed outline of the modified contact approach is shown on the next page.

	New Address Available	No Address Available
Step 1	<p>A paper questionnaire will be mailed to the targeted respondent at the new address. The paper-and-pencil mail survey will include a postage-paid return envelope. The survey will include a series of frequently asked questions that will give information about the survey and assurances that there is no risk associated with participation.</p> <p>Respondents will also be informed of the option to complete a web survey. This mailing will also include the letter from an IRS official endorsing the survey and emphasizing the importance of the data collection effort.</p> <p>Because of time constraints, this package will be sent via UPS.</p>	<p>If a phone number is available for the respondent, an interviewer will contact the respondent to request an updated mailing address. The interviewer will briefly introduce the survey and will be able to answer questions about the survey and the process. In addition, the interviewer will administer the interview over the telephone if the respondent wishes.</p>
Step 2	<p>If no completed survey is received, respondents will receive a reminder prompt from a telephone interviewer asking respondents to complete the survey. Telephone interviewers will be prepared at this stage to administer the interview over the telephone if the respondent wishes.</p> <p>If no phone number is available, respondents will receive a reminder postcard asking the respondent to please complete the survey.</p>	<p>If a respondent is willing to provide an updated mailing address, a survey package will be mailed to the respondent at the new address. The paper-and-pencil mail survey will include a postage-paid return envelope. The survey will include a series of frequently asked questions that will give information about the survey and assurances that there is no risk associated with participation. Respondents will also be informed of the option to complete a web survey. This mailing will also include the letter from an IRS official endorsing the survey and emphasizing the importance of the data collection effort.</p>
Step 3	N/A	<p>If no completed survey is received, respondents will receive a reminder prompt from the same telephone interviewer who contacted them in Step 1. This interviewer will be prepared to administer the interview over the telephone if the respondent wishes.</p>

The secure web survey will be posted online using a proprietary web survey delivery system developed by our contractor, Westat. The software easily accommodates different question formats, including open-ended response fields. It also allows participants to skip questions and complete the survey in more than one session (i.e., the respondent can leave the web survey and come back to finish it at a later time). Development and testing of the web survey will follow well-established, documented best methods.

The paper-and-pencil mail survey will be designed to be user friendly, easy to navigate, and with clear and simple instructions. The survey will be created using TeleForm technology, a software system for intelligent data capture and image processing. The software extracts indexing information automatically from any document type through the use of multiple recognition engines. TeleForm reads hand print, machine print, optical marks, bar codes, and signatures.

Response data will be stored and tracked in a response database which can then used to update and extend the IRS

compliance burden model. In addition, a tailored Survey Management System will track cases throughout all modes of contact, including mail, telephone, and IVR.

3. Describe methods to maximize response rates and to deal with issues of non-response.

Upon completion of the survey protocol, we will conduct a non-response bias analysis. This analysis will be the same as what was done for previously-conducted surveys, using a raking technique as a way to control for bias in a multivariate scenario. The process is further outlined in the paper “Response Mode and Bias Analysis in the IRS’ Individual Taxpayer Burden Survey”, by J. Michael Brick, George Contos, Karen Masken, and Roy Nord.

4. Describe any tests of procedures or methods to be undertaken.

To ensure that the collection of information is not burdensome and that the questions are clearly written and will produce accurate and valid results, the IRS will conduct cognitive testing for any new or revised survey instrument. Cognitive testing is a well-established qualitative research method intended to identify problems respondents have with comprehension of survey questions (Willis 2005)¹. The testing will be conducted with taxpayers in the Washington, D.C. area. Respondents will be recruited according to specific criteria (e.g., filing status, complexity of return, and filing method). Efforts will be made to recruit respondents who are demographically representative of the population being surveyed.

In addition, at the outset as well as after each interaction of testing, the instrument will undergo extensive review by the IRS, the contractor, and stakeholders.

5. Provide the names and telephone numbers of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

IRS Office of Research

Statistical Design:

Michael Sebastiani, 202-874-0831

Wei Liu, 202-874-0575

Karen Masken, 202-874-4321

Leann Weyl, 202-874-0559

Collection and Analysis:

John Guyton, IRS Research, Analysis & Statistics

George Contos, IRS Research, Analysis & Statistics

Ronald H. Hodge II, IRS Research, Analysis & Statistics

Sandy Lin, IRS Research, Analysis & Statistics

Brenda Schafer, IRS Research, Analysis & Statistics

Melissa Vigil, Research, Analysis & Statistics

Edith Brasher, Department of Treasury, Office of Tax Policy

Allen Lerman, Department of Treasury, Office of Tax Policy

Susan Nelson, Department of Treasury, Office of Tax Policy

Gerald Silverstein, Department of Treasury, Office of Tax Policy

Westat

Data Collection:

Jocelyn Newsome, Research Analyst, 301-212-3734

Kerry Levin, Project Manager

Jennifer O’Brien, Project Director

Martha Stapleton, Project Manager

¹Willis, G.B. (2005). Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA: Sage Publications.

Statistics:
Mike Brick, Statistician