

General Instructions

A Supporting Statement, including the text of the notice to the public required by 5 CFR 1320.5(a)(i)(iv) and its actual or estimated date of publication in the Federal Register, must accompany each request for approval of a collection of information. The Supporting Statement must be prepared in the format described below, and must contain the information specified in Section A below. If an item is not applicable, provide a brief explanation. When Item 17 of the OMB Form 83-I is checked "Yes", Section B of the Supporting Statement must be completed. OMB reserves the right to require the submission of additional information with respect to any request for approval.

Specific Instructions

A. Justification

1. Explain the circumstances that make the collection of information necessary. Identify any legal or administrative requirements that necessitate the collection. Attach a copy of the appropriate section of each statute and regulation mandating or authorizing the collection of information.

Each year, individuals and businesses in the United States submit more than 200 million tax returns to the Internal Revenue Service (IRS). The IRS uses the information in these returns, recorded on hundreds of distinct forms, to administer a tax system whose rules span thousands of pages. Managing such a complex and broad-based tax system is costly but represents only a fraction of the total burden of the tax system. Equally, if not more burdensome, is the time and out-of-pocket expenses that citizens spend in order to comply with tax laws and regulations.

To understand the full measure of this burden, the IRS contracted IBM Global Business Services (IBM) to measure and model taxpayer burden for Wage and Investment (W&I) and Self-Employed (SE) taxpayers. The approach includes administering a survey that addresses the time and money taxpayers spend as a result of fulfilling their federal income tax obligations. IRS and IBM conducted initial surveys with the W&I and SE populations in 1999 and 2000, respectively. Since that time, changes in tax regulations, tax administration, and taxpayer behavior have significantly altered the amount and distribution of taxpayer burden. Therefore, we have designed a significantly revised survey that will better reflect the current tax rules and regulations, the increased usage of tax preparation software, the increased use of filing returns electronically, and other information collection needs.

Updated information from this survey is necessary for several reasons. It would help IRS:

- Better understand taxpayer time and out-of-pocket burden
- Gain insights into the impact of tax law and tax administration changes on taxpayer burden
- Enhance the accuracy of the information collection budget estimates it provides under the Paperwork Reduction Act (44 U.S.C. 3501 et seq.)
- Fulfill its mission to provide top quality service to taxpayers

There are no other legal or administrative requirements that necessitate the collection.

2. Indicate how, by whom, and for what purpose the information is to be used. Except for a new collection, indicate the actual use the agency has made of the information received from the current collection.

The IRS is working with IBM to develop an improved method for measuring, estimating, and modeling taxpayer burden. The data collected from this survey of individual taxpayers will be used as an input to a micro-simulation model that measures taxpayer burden. IRS also will publish the updated burden estimates in tax form instructions to inform taxpayers. Data will be collected by IBM using a mixed mode

(i.e., telephone, mail, and web-based) data collection methodology. The primary method of collecting the survey data will be via a telephone interview. Taxpayers for whom a working telephone number is not available or who are not reached via telephone will be able to complete a self-administered mail or web-based questionnaire.

The information collected via this study will be used by IRS to achieve several important goals:

- Measure individual taxpayer burden more comprehensively and accurately
- Provide policymakers with a tool to help guide efforts to reduce taxpayer burden
- Allow IRS to determine the effectiveness of its programs in reducing taxpayer burden
- Provide policymakers with a tool to explain taxpayer burden
- Complement existing customer satisfaction measures

3. Describe whether, and to what extent, the collection of information involves the use of automated, electronic, mechanical, or other technological collection techniques or other forms of information technology, e.g., permitting electronic submission of responses, and the basis for the decision for adopting this means of collection. Also describe any consideration of using information technology to reduce burden.

The best available methods for collecting the necessary data while minimizing the burden on the public will be employed. IBM will use state-of-the-art methods and interviewing techniques currently employed in the survey research industry to elicit cooperation from potential respondents and thus maximize response. Implicit in these methods is a minimization of respondents' burden, both real and perceived, associated with survey participation. IBM will use its CATI (Computer Assisted Telephone Interviewing) system to collect the information for the telephone interview portion of the data collection methodology. It will also provide web-based survey and mail survey options where appropriate.

4. Describe efforts to identify duplication. Show specifically why any similar information already available cannot be used or modified for use for the purposes described in Item 2 above.

The only way to obtain information about the amount of time and out-of-pocket expenses associated with complying with the U.S. tax system is to ask taxpayers directly. Before developing the survey instrument, attempts were made to eliminate duplication of information available within the agency wherever possible. This collection request reflects the fact that a substantial amount of the data required to update the model to measure burden is available from the IRS. Only those items that are not available in IRS administration databases (e.g., measures of time and out-of-pocket burden) are included in the survey instrument.

5. If the collection of information impacts small businesses or other small entities (Item 5 of OMB Form 83-I), describe any methods used to minimize burden.

This collection focuses on individual taxpayers (both Wage & Investment and Self-Employed) and thus, does not impact small businesses or other small entities as defined in Form 83-I.

6. Describe the consequence to Federal program or policy activities if the collection is not conducted or is conducted less frequently, as well as any technical or legal obstacles to reducing burden.

If this survey is not conducted, the IRS will continue to base its burden estimates on survey data that is more than 8 years old and does not reflect current administrative and legislative tax policies. It also does not reflect rather dramatic changes in information technology that have impacted IRS services and taxpayer behavior. In light of these changes, there is a strong belief that the taxpayer burden model needs to include updated taxpayer information. In addition, if this study is not conducted, it will limit the IRS's ability to provide accurate information for the IRS Information Collection Budget for OMB.

7. Explain any special circumstances that would cause an information collection to be conducted in a manner:

- requiring respondents to report information to the agency more often than quarterly;
- requiring respondents to prepare a written response to a collection of information in fewer than 30 days after receipt of it;
- requiring respondents to submit more than an original and two copies of any document;
- requiring respondents to retain records, other than health, medical, government contract, grant-in-aid, or tax records, for more than three years;
- in connection with a statistical survey, that is not designed to produce valid and reliable results that can be generalized to the universe of study;
- requiring the use of a statistical data classification that has not been reviewed and approved by OMB;
- that includes a pledge of confidentiality that is not supported by authority established in statute or regulation, that is not supported by disclosure and data security policies that are consistent with the pledge, or which unnecessarily impedes sharing of data with other agencies for compatible confidential use; or
- requiring respondents to submit proprietary trade secrets, or other confidential information unless the agency can demonstrate that it has instituted procedures to protect the information's confidentiality to the extent permitted by law.

Data for this survey will be collected only once from any individual. Confidentiality is addressed in Section A-10 of this package.

Survey respondents for this information collection will:

- have more than 30 days to respond to the mail questionnaire (except at the end of the data collection period),
- not be asked to submit any documents,
- not be asked to retain any records, and
- not be asked to provide any proprietary trade secrets or confidential information.

The sample design will allow the survey results to be generalized to the individual taxpayer population, and IBM will use only statistical data classifications that have been reviewed and approved by OMB.

8. If applicable, provide a copy and identify the date and page number of publication in the Federal Register of the agency's notice, required by 5 CFR 1320.8(d), soliciting comments on the information collection prior to submission to OMB. Summarize public comments received in response to that notice and describe actions taken by the agency in response to these comments.

- Specifically address comments received on cost and hour burden.
- Describe efforts to consult with persons outside the agency to obtain their views on the availability of data, frequency of collection, the clarity of instructions and recordkeeping, disclosure, or reporting format (if any), and on the data elements to be recorded, disclosed, or reported.
- Consultation with representatives of those from whom information is to be obtained or those who must compile records should occur at least once every 3 years - even if the collection of information activity is the same as in prior periods. There may be circumstances that may preclude consultation in a specific situation. These circumstances should be explained.

Although the project is managed by IRS, there is a working group that provides input to the project at key decision points. The non-IRS members of the working group include:

- Allen Lerman (Department of the Treasury, Office of Tax Analysis)
- Susan Nelson (Department of the Treasury, Office of Tax Analysis)

- Janet Hotzblatt (Department of the Treasury)
- Joe Cordes (George Washington University)
- Tom Beers (National Tax Advocate)

The following stakeholders, outside of the IRS Project Management Team, were consulted:

- Mark Mazur (Internal Revenue Service)
- Janice Hedemann (Internal Revenue Service)
- Ronnie Moore (Internal Revenue Service)
- Jodi Patterson (Internal Revenue Service)
- Nancy Farrell (Internal Revenue Service)

This is a one-time data collection effort; periodic data collection cycles are not envisioned.

9. Explain any decision to provide any payment or gift to respondents, other than re-enumeration of contractors or grantees.

The use of incentives in survey research is a way to help increase the response rate, especially in voluntary studies. Incentives may be given prior to the survey taking place regardless of participation, or retained as a reward for those who complete the survey. They may take the form of a monetary or non-monetary gift. The decision to offer an incentive encourages responses and demonstrates to respondents that their time is appreciated (Simmons and Wilmot, 2004)¹.

Studies of incentives examine a variety of methods; however, the research can generally be classified into one of four types by incentive type and point of payment as shown in the following table.

	Cash Incentives	Non-Cash Incentives
Pre-paid (Provided Prior to the Interview)	This incentive type refers to monetary payments, made either by cash or check, presented before the actual interview is completed.	Non-cash incentives typically take the form of gift items, such as stamps, calculators, etc., and are provided to the respondent in advance of the survey interview.
Promised (Presented Upon Completion of the Interview)	This type of monetary incentive is promised to the survey respondent before the interview, but is presented only after he or she participates in the survey as a reward for the interview.	Promised non-cash incentives are presented to the respondent either immediately after the interview or a short time later. Popular forms of this incentive include survey publications, sweepstakes entries, and donations to charity.

According to the published literature on survey incentives, pre-paid cash incentives have consistently generated the greatest number of successes and demonstrated the largest increases in response rate. Non-cash incentives, when pre-paid, have also shown a number of significant increases. While pre-paid monetary incentives have been successful for generating higher response rates, it is also interesting to note that even a very small amount of money, such as a one or five-dollar bill, can significantly increase response rates. In fact, small monetary gifts appear to be as successful, or more successful, than larger quantities of money.

¹ Simmons E & Wilmot A (2004). Incentive payments on social surveys: A literature review. *Survey Methodology Bulletin*, No. 53, 1-11.

For this data collection effort, the IRS study team proposes utilizing a combination of pre-paid and promised incentives. All potential respondents will receive an advance letter, which will provide a detailed description of the purpose of the survey. With this letter, a one dollar bill will be included as "an attention getter." The letter will also indicate that respondents, who successfully complete the questionnaire by phone, mail, or web, will be sent an incentive of about \$25.00. The promised \$25.00 incentive will be sent to qualifying respondents shortly afterward.

10. Describe any assurance of confidentiality provided to respondents and the basis for the assurance in statute, regulation, or agency policy.

Strict confidentiality will not be promised to respondents, since IRS intends to access individual survey responses as part of its on-going burden research. Therefore, the questionnaire states explicitly that responses will be seen by IRS research staff; however, respondents are assured that their information will be used only for research on taxpayer burden. The questionnaire will also explain that responses will not be used for any current or future enforcement activities.

11. Provide additional justification for any questions of a sensitive nature, such as sexual behavior and attitudes, religious beliefs, and other matters that are commonly considered private. This justification should include the reasons why the agency considers the questions necessary, the specific uses to be made of the information, the explanation to be given to persons from whom the information is requested, and any steps to be taken to obtain their consent.

The survey itself does not include questions that would commonly be considered private or sensitive in nature. Ultimately, IRS research staff will match the respondents' survey data to their administrative information (contained in IRS data sources), in order to help approximate time and money burden for the tax filing population. While certain fields on the administrative data could be considered sensitive (e.g., salary), respondents will not be asked to provide this information via the survey.

12. Provide estimates of the hour burden of the collection of information. The statement should:

- Indicate the number of respondents, frequency of response, annual hour burden, and an explanation of how the burden was estimated. Unless directed to do so, agencies should not conduct special surveys to obtain information on which to base hour burden estimates. Consultation with a sample (fewer than 10) of potential respondents is desirable. If the hour burden on respondents is expected to vary widely because of differences in activity, size, or complexity, show the range of estimated hour burden, and explain the reasons for the variance. Generally, estimates should not include burden hours for customary and usual business practices.
- If this request for approval covers more than one form, provide separate hour burden estimates for each form and aggregate the hour burdens in Item 13 of OMB Form 83-I.
- Provide estimates of annualized cost to respondents for the hour burdens for collections of information, identifying and using appropriate wage rate categories. The cost of contracting out or paying outside parties for information collection activities should not be included here. Instead, this cost should be included in Item 13.

The total number of potential respondents will be approximately 15,000. Each of them will receive an advance letter describing the purpose of the study, which they may spend about one minute reading. Of the 15,000 potential respondents, approximately 10,000 will complete the entire questionnaire and 5,000 will screen out because they do not meet the study requirements. Each potential respondent will participate only once. We estimate it will take approximately 5 minutes for a potential respondent to refuse to participate or screen out. We also expect a completed interview to take approximately 20 minutes. We estimate that it will take approximately the same time to complete the phone, mail, and web versions of the questionnaire. The content included in each instrument will be the same. These time burden estimates are based on the previous surveys conducted in 1999 and 2000. Since the new survey

has been designed with a similar length, these estimates should be comparable. We, therefore, estimate that the main portion of the study will take approximately 4,000 hours (itemized in the table below).

Activity (Main Study)	Number of Respondents	Frequency of Response	Average Time (In Minutes)	Total Time (In Hours)
Read Advance Letter	15,000	1	1	250
Refuse to Participate or Screen Out	5,000	1	5	417
Complete Entire Questionnaire	10,000	1	20	3,333
Total Time Burden (In Hours)				4,000

A pretest with approximately 225 potential respondents also will be conducted. We estimate that there will be approximately 75 screen-outs and 150 completed interviews. The average time and frequency will be the same as described above. We, therefore, estimate that the pretest will take approximately 60 hours (itemized in the table below).

Activity (Pre-Test)	Number of Respondents	Frequency of Response	Average Time (In Minutes)	Total Time (In Hours)
Read Advance Letter	225	1	1	4
Refuse to Participate or Screen Out	75	1	5	6
Complete Entire Questionnaire	150	1	20	50
Total Time Burden (In Hours)				60

In terms of the annualized cost to respondents for the hour burdens for the information collection, we assume a wage rate of \$16.76, which represents the BLS total private average hourly wage rate from 2006. Therefore, the estimated cost burden of the main study would be \$67,040.00 ((250 + 417 + 3,333) * \$16.76). For the pre-test, the cost would be \$1,005.60 ((4 + 6 + 50) * \$16.76). In total, the estimated cost burden would be \$68,045.60.

13. Provide an estimate for the total annual cost burden to respondents or recordkeepers resulting from the collection of information. (Do not include the cost of any hour burden shown in Items 12 and 14).

- The cost estimate should be split into two components: (a) a total capital and start-up cost component (annualized over its expected useful life) and (b) a total operation and maintenance and purchase of services component. The estimates should take into account costs associated with generating, maintaining, and disclosing or providing the information. Include descriptions of methods used to estimate major cost factors including system and technology acquisition, expected useful life of capital equipment, the discount rate(s), and the time period over which costs will be incurred. Capital and start-up costs include, among other items, preparations for collecting information such as purchasing computers and software; monitoring, sampling, drilling and testing equipment; and record storage facilities.
- If cost estimates are expected to vary widely, agencies should present ranges of cost burdens and explain the reasons for the variance. The cost of purchasing or contracting out information collections services should be a part of this cost burden estimate. In developing cost burden estimates, agencies may consult with a sample of respondents (fewer than 10), utilize the 60-day pre-OMB submission public comment process and use existing economic or regulatory impact analysis associated with the rulemaking containing the information collection, as appropriate.
- Generally, estimates should not include purchases of equipment or services, or portions thereof, made: (1) prior to October 1, 1995, (2) to achieve regulatory compliance with requirements not associated with the information collection, (3) for reasons other than to provide information or keep records for the government, or (4) as part of customary and usual business or private practices.

There will be no direct cost to respondents for their participation in this survey. All respondents will be contacted by IBM through the phone, mail, and web. Mail surveys will be given stamped envelopes to return their responses.

14. Provide estimates of annualized costs to the Federal government. Also, provide a description of the method used to estimate cost, which should include quantification of hours, operational expenses (such as equipment, overhead, printing, and support staff), and any other expense that would not have been incurred without this collection of information. Agencies may also aggregate cost estimates from Items 12, 13, and 14 in a single table.

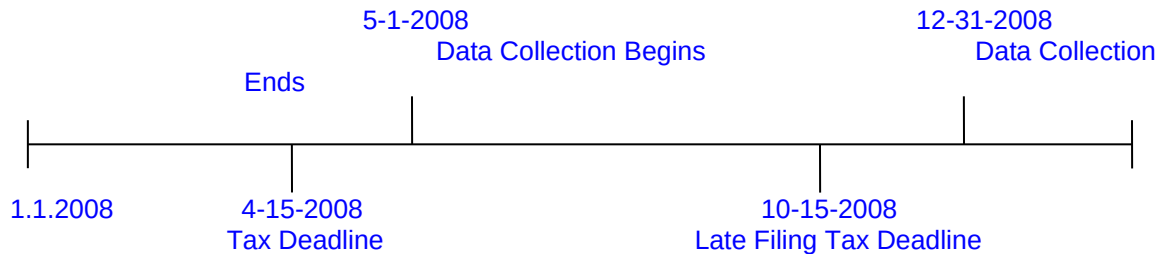
The one-time cost of Items 12 and 13 and the one-time operational cost of conducting the survey (the estimated total contract value to the contractor who will conduct the survey) are summed below.

$\$68,045.60 + \$0.00 + \$2,000,000 = \$2,068,045.60$

15. Explain the reasons for any program changes or adjustments reported in Items 13 or 14 of the OMB Form 83-I.

Not applicable. There are no program changes or adjustments.

16. For collections of information whose results will be published, outline plans for tabulation and publication. Address any complex analytical techniques that will be used. Provide the time schedule for the entire project, including beginning and ending dates of the collection of information, completion of report, publication dates, and other actions.



The data collection period will begin soon after the regular tax filing ends on April 15, 2008 (e.g., late April or early May 2008). Collection will continue through the late filing deadline of October 15, 2008 until December 2008. Ongoing progress reports will be submitted throughout the survey effort ending 60 days after the conclusion of data collection.

Initial results of the survey will be presented in a paper at the IRS research conference in June 2009. Detailed results from the survey will be used to update the estimated relationships between taxpayer burden and taxpayer filing attributes. These estimates will then be used to update the ITBM (Individual Taxpayer Burden Model), a micro-simulation model that provides taxpayer burden estimates and other tax-related information. In addition, the ITBM allows the IRS to estimate the changes in taxpayer burden resulting from changes in tax law and/or administrative regulations. IRS will use the results of the model to develop annualized burden estimates, which are necessary for compliance with the information collection budget associated with all IRS forms (OMB No. 1545-007).

17. If seeking approval to not display the expiration date for OMB approval of the information collection, explain the reasons that display would be inappropriate.

The required OMB approval information, as well as the expiration date for the OMB approval, will be displayed on all survey documents and read to telephone interview respondents.

18. Explain each exception to the certification statement identified in Item 19, "Certification for Paperwork Reduction Act Submissions," of OMB Form 83-I.

Not applicable. No exceptions are believed to exist.

B. Collections of Information Employing Statistical Methods

The agency should be prepared to justify its decision not to use statistical methods in any case where such methods might reduce burden or improve accuracy of results. When Item 17 on the Form OMB 83-1 is checked, "Yes," the following documentation should be included in the Supporting Statement to the extent that it applies to the methods proposed:

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

The potential respondent universe is comprised of Wage & Investment and Self Employed taxpayers living in the United States. These taxpayers file a Form 1040, Form 1040A, or Form 1040EZ (as well as other forms and schedules). For this survey, we are focusing on Tax Year 2007. The sample frame will be developed using IRS administrative data sources, such as the Returns Transaction File (RTF).

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.
- **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/her return.
 - o Prepared by a Paid Professional (Paid)
 - o Prepared using Tax Preparation Software (Soft)
 - o Prepared by Hand (Self)
- **Differential burden.** Variable consisting of concepts, such as taxpayer characteristics and relevance of tax provisions, which contribute to a taxpayer's tax situation being burdensome.
 - o Low
 - o Low-Medium
 - o Medium
 - o Medium-High
 - o High

The differential burden variable 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 are included. The tax preparation method variable ensures both a proper balance and an adequate representation of paid preparers, software preparers, and self preparers.

For each stratification variable, categories were chosen to increase the precision of sample estimates. Specifically, we define categories that minimize burden variance within strata and maximize it between strata. As a result, sample size can then be concentrated on taxpayer segments with heterogeneous burden, thus increasing the precision of population estimates.

Considerations in determining the selection probability for each stratum include the number of taxpayers in the stratum, the diversity of taxpayers in the stratum, and interest in the elements within the stratum as separate subject analysis group. Sampling rates are increased in strata where the variance among the elements is large at the expense of sampling rates in strata with smaller element variances. This method of stratification is referred to as optimum allocation.

The table below illustrates the optimum allocation method in practice for monetized burden assuming an overall sample size of 10,000. Since Tax Year 2006 data is not available yet, we used the 2005 Continuous Work History Sample (CWHS) dataset to estimate the following population counts for the proposed strata definition. We used the previously collected survey data to estimate the mean burden and standard deviations for each stratum. With these inputs, we were able to use the Neyman Allocation Method to determine the sample sizes for the strata. Since this method works for one dependent variable at a time – and we have two: time and money – we created a monetized burden variable by multiplying the total time burden by \$16.13 and adding it to the total money burden. The \$16.13 is the BLS total private average hourly wage rate from 2005.

Total Monetized Burden	
Neyman Allocation Overall CV =	1.26%

Monetized Burden Strata	CWHS (TY05)		Survey (TY99/TY00)		Neyman Allocation	
	N	%	Mean	Std Dev	n	CV
Paid, Low	10,944,000	8.17%	190.75	384.28	433	9.68%
Paid, Low-Medium	24,250,000	18.11%	258.03	569.25	1,422	5.85%
Paid, Medium	17,928,000	13.39%	451.01	722.50	1,334	4.39%
Paid, Medium-High	15,998,000	11.95%	975.05	974.14	1,605	2.49%
Paid, High	10,944,000	8.17%	2148.50	2143.75	2,417	2.03%
Self, Low	11,248,000	8.40%	168.05	517.76	600	12.58%
Self, Low-Medium	9,154,000	6.84%	261.23	507.05	478	8.88%
Self, Medium	6,706,000	5.01%	366.49	337.65	233	6.04%
Self, Medium-High	2,880,000	2.15%	785.69	631.62	187	5.88%
Self, High	904,000	0.68%	1120.01	695.25	65	7.70%
Soft, Low	5,664,000	4.23%	202.11	425.51	248	13.37%
Soft, Low-Medium	6,828,000	5.10%	291.72	464.06	326	8.81%
Soft, Medium	6,022,000	4.50%	497.70	442.47	274	5.37%
Soft, Medium-High	3,242,000	2.42%	1005.07	755.34	252	4.73%
Soft, High	1,184,000	0.88%	1670.44	1015.03	124	5.46%
Total	133,896,000	100.00%	574.53		9,998	

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 cases to meet the needs of the modeling tool to identify the determinants of burden. The second 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 range requirements).

The final survey sample will be selected from the IRS administrative records of tax forms filed in Tax Year 2006 (Processing Year 2007). Operationally, the sampling units may be selected into each stratum leveraging the “transform taxpayer identification number” technique used by IRS for selecting samples from large data files.

The response rates are the proportion of completed interviews in the total number of eligible respondents.

The overall response rate for the previous W&I study was 60.5% while the one for the previous SE study was 56.4%. The targeted response rate is 70 percent, and IBM will employ methods described in Section B-3 (Methods to Minimize Non-Response) to achieve the highest possible response rate.

2. Describe the procedures for the collection of information including:

- Statistical methodology for stratification and sample selection,
 - Estimation procedure,
 - Degree of accuracy needed for the purpose described in the justification,
 - Unusual problems requiring specialized sampling procedures, and
 - Any use of periodic (less frequent than annual) data collection cycles to reduce burden.
-

The statistical methodology for stratification and sample selection is presented in Section B-1.

Sampling weights will be used to generalize survey results from the sample to the population. They are needed because sampling rates differ between strata (and are fixed within strata). Unbiased population estimates require that the sampling weight for each observation be equal to the inverse of the probability that the observation was selected into the sample.

Virtually all survey data are weighted before they can be used to produce reliable estimates of population parameters. While reflecting the selection probabilities of sampled units, weighting also attempts to compensate for practical limitations of a sample survey, such as differential non-response and undercoverage. Furthermore, by taking advantage of auxiliary information about the target population, weighting can reduce the variability of estimates. The weighting process essentially entails four major steps:

- Computing the design or base weights
- Adjusting the base weights for non-response
- Adjusting the non-response-adjusted weights for frame inadequacies such as undercoverage
- Processing the adjusted weights through a series of quality control checks to detect extreme outliers and to prevent any computational as well as procedural errors

Since the IRS study team has a considerable amount of information about this population, all respondents will be weighted based on the demographic makeup of the population of individual taxpayers. National estimates based directly on results from this survey will have a CV of less than 2%.

Total Monetized Burden	
Neyman Allocation Overall CV =	1.26%

No unusual problems requiring specialized sampling procedures are anticipated.

This is a one-time data collection effort; periodic data collection cycles will not be employed.

3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

For the survey, while a response rate of 70 percent is targeted, every attempt will be made to achieve the highest possible response rate. Several steps will be taken to maximize response rates:

- **Advance Letters.** Potential respondents will receive an advance letter, which will explain the purpose of the survey and urge their participation.
- **Incentives.** A one-dollar "attention getter" will be attached to the advance letter, and a \$25.00 cash incentive will be provided to respondents who complete the questionnaire.
- **Assurances.** Respondents will be informed that their responses will be used for burden research only. Their responses will not be used for current or future enforcement activities.
- **Multiple Data Collection Modes.** IBM will offer phone, mail, and web-based data collection modes.
- **Rigorous Follow-up.** IBM will employ a rigorous follow-up protocol to obtain a completed interview.

Non-respondent follow-up will include web/mail and telephone modes. Procedures and steps aimed at maximizing the response rate from telephone interviewing are as follows:

1. Using skillful, experienced and persuasive interviewers. Most refusals to interview occur before the interview begins; well-prepared interviewers can work through this stage most proficiently. Interviewer training activities will emphasize this critical activity.
2. Allowing participants to confirm the legitimacy of the survey effort. Many potential respondents need to be convinced that they will be taking part in a valid, legally authorized research project before consenting to be interviewed. If asked, IBM will provide the address of the OMB and IRS departments where potential respondents can write to confirm the credibility of the study, and to answer questions concerning the survey.
3. Having the interviewer call attention to the lack of risks or costs associated with participation. Respondents will be assured that all responses will be use only for research.
4. Giving the respondents the opportunity to refuse answers to individual question. Often, potential respondent's concerns are deflected or minimized when they are told they can choose to answer questions selectively. Although this technique seems to favor reducing overall non-response by increasing some item non-response, experience show that respondents rarely refuse to answer more questions when presented with this offer than they would otherwise.
5. Making the interview effort as convenient as possible. Callbacks at more convenient times will be encouraged, as will appointment setting and completion of the interview in multiple interview sessions if necessary. Flexibility geared to the respondent's needs helps avoid an outright refusal.
6. Providing a substantial schedule of attempts. At least ten or more attempts will be made by telephone to reach each sampled taxpayer. If the final call results in a scheduled callback, an additional attempt will be made to obtain a completed interview by mail or computer.
7. Varying the time of callbacks. Initial calls will be concentrated during business hours. When callbacks are necessary, they will occur on a different day and at a different time. Contacts on travel or otherwise away, will be scheduled for callbacks upon their return.

8. Distinguishing “hard refusals” from “reluctant to respond.” The non-response results on the call report will direct efforts in follow-up contacts. Refusals will be dismissed from the sample only after individual examination has determined that they are “hard” refusals and there is no feasible alternative.
9. Respondents will receive a letter informing them of the study. The letter will explain the purpose of the study and invite them to respond.
10. Respondents will be able to contact IBM via a toll-free number. When necessary, respondents may call a toll-free number to schedule or conduct the interview.

The above procedures will be adjusted as necessary to solicit maximum participation from respondents. Some potential respondents will never be reached during the telephone interview phase (e.g., they were never at home), and some who were reached will not respond to the telephone contact. In addition, telephone contact with some potential respondents will not be possible due to an inability to locate a correct phone number. For each of these instances, the supplemental mail or web survey methodology will be employed.

If requested, a mail or web-based questionnaire will be sent to respondents. These self-administered questionnaires will cover the same questions as the telephone survey questionnaire. They will provide supporting explanations, a glossary of terms, and other information necessary for the respondent to complete the questionnaire. The mail package will also include postage paid return envelopes. Up to three mailings will be made to each potential respondent.

Like most sample surveys, this taxpayer survey will experience unit nonresponse despite efforts to maximize response rates. Survey nonresponse is usually compensated for by some form of weighting adjustment to reduce the bias in survey estimates. Sometime, a weighting class nonresponse adjustment, using socio-economic and demographic variables to create the weighting classes, is used to adjust for potential nonresponse bias. An alternative method for forming nonresponse adjustment cells is to use response propensities. Nearest neighbor matching or multiple imputation can also be used for item non-response. These techniques preserve the distribution of item values, permit the use of the same sample weight for all items, and enable results obtained from different analyses to be consistent with one another.

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of test may be submitted for approval separately or in combination with the main collection of information.

The telephone questionnaire, mail questionnaire, and web questionnaire will be tested thoroughly prior to survey administration. Question order, understandability, and readability will be evaluated. A pre-test of the programs, data management, and survey administration processes will be performed. All refinements to the instruments and processes based on the results of these pre-tests will be undertaken to improve the questionnaires. We anticipate doing 150 pre-tests before we begin data collection.

5. Provide the name and telephone number 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.

IBM Global Business Services will collect and analyze the information for the agency. The following individuals were consulted on the statistical aspects of the design:

Dr. Donald DeLuca
IBM Global Business Services
6710 Rockledge Drive
Bethesda, MD 20817
301-803-1332
don.deluca@us.ibm.com

Dr. Arnold Greenland
IBM Global Business Services
6710 Rockledge Drive
Bethesda, MD 20817
301-803-6746
agreenland@us.ibm.com

Mr. Derek Snidauf
IBM Global Business Services
71 South Wacker Drive
Chicago, IL 60606
312-529-1620
derek.snidauf@us.ibm.com