# Measuring Nonsampling Error in the Statistics of Income Individual Tax Return Study

*Jana Scali, Valerie Testa, Maureen Kahr, and Michael Strudler, Internal Revenue Service*

Data collection for the SOI Individual Study begins with a sample of administrative tax records. While the sample is being transcribed, small subsamples of returns are randomly chosen and independently transcribed and processed for a quality evaluation. The IRS Statistics of Income (SOI) Division has an Individual Systematic Improvement (ISI) System which is the tool used to create the quality review sample and improve the Individual Tax Return Study data. The purpose of this paper is to estimate a component of nonsampling error in the SOI Individual Study. The data from the quality review process is used for this purpose.

The paper is organized as follows. We describe SOI's Individual sample design along with some sources of nonsampling error. We describe the editing process and the Individual Systematic Improvement (ISI) System used by SOI to evaluate and improve the quality of the Individual 1040 Program. We describe the study and its limitations. We explain the model used to estimate nonsampling error. We show the Index of Inconsistency. We cover the Intra-Editor Correlation Coefficient and Design Effect by element followed by conclusions.

## ▶ Individual Sample and Nonsampling Error Description

The statistics for the SOI Individual Study are estimates from a probability sample of unaudited Individual Income Tax Returns filed by U.S. citizens and residents during Calendar Year 2004. The estimates represent all returns filed for Tax Year 2003 with a small number representing prior years. For Tax Year 2003, some 184,988 returns were sampled from a population of 131,291,334.

The sample consists of two parts. The first part is a stratified probability sample, in which the population of tax returns is classified into subpopulations, called strata, and a sample is randomly selected independently from each stratum. Strata are defined by the type of return submitted by the taxpayer. A Bernoulli sample is independently selected from each stratum with rates ranging from .05 percent to 100 percent. The second part of the sample is a random sample based on the primary taxpayer's Social Security number. If the last four digits of the primary taxpayer's Social Security number listed on the tax return equals one of five predetermined endings, then the tax return is included in the sample.

The quality of a sample estimator is a function of both sampling and nonsampling errors. Sampling errors arise due to drawing a probability sample rather than conducting a census. Nonsampling errors are due to data collection and processing procedures. They can be the result of misleading definitions and concepts or defective methods of data collection, tabulation, and coding. Nonsampling errors may increase with sample size, and, if not properly controlled, they can be more damaging to a study than sampling errors.

There are four components of nonsampling error. Coverage or frame errors occur when someone does not file a tax return. Nonresponse errors (missing data) arise when the Statistics of Income Division is unable to obtain the tax return because another function within the Internal Revenue Service has the return. Measurement errors are differences in the reported and the actual values. These errors are taxpayer errors. Processing errors occur at the data processing stage. They include editing, coding, data entry, and programming errors. This paper will describe and measure processing errors, which arise due to the following factors:

1. Lack of trained and experienced editors including quality supervisors.

2. Errors in data processing operations such as coding, keying, verification, and tabulation.

3. Procedural, Systemic, or Organizational Defects such as improper instructions, in-

adequate training, and insufficient time to complete a return.

Nonsampling errors are very important to measure because they can cause large biases and produce unreliable estimates if not controlled. By following the correct procedures during sample selection through the analysis of results, nonsampling errors can be controlled and dramatically decreased.

## ▶ SOI Editing and Quality Review Processes

For SOI purposes, when we mention editing, it refers to the process of an individual transcribing data items or elements from the tax return into our database. An element is a specific line item from a tax return. The individual transcribing the data is referred to as an editor. For the SOI Individual Study, 97 editors at four IRS Submission Processing Centers edited data from Individual income tax returns selected for the 2003 SOI sample. The data extracted come from Forms 1040, 1040A, and 1040EZ individual income tax returns and approximately 45 associated forms and schedules.

To assist the editors in this process, SOI's National Office analysts in Washington, DC, implement various procedures to make the edited data adhere to individual tax standards and to try to keep the editing process as consistent as possible across the four centers. For example, the editors receive extensive training on the data editing process and correction procedures before they begin editing individual tax return data for the SOI sample. Then, as data are edited, numerous computerized tests are performed on the extracted data to ensure that certain accounting conditions are satisfied and that data are consistent across forms. All of these computerized tests are reviewed and tested by National Office staff prior to data extraction in a process called Systems Acceptability Testing. Various utilities and help features to aid in the edit process are also built into the computer edit system. For instance, there are utilities that list valid codes and definitions for a particular item. In addition, there is a feature that allows data from the previous year's tax return to be viewed. There is also a comprehensive editing manual that contains detailed instructions and procedures that editors are expected to follow while transcribing and correcting the tax return data. The editing manual for the 2003 sample was just over 600 pages.

During data editing, a simple random sample of one or two returns each week is selected for each editor for regular quality review. The goal is to have approximately 50 returns per editor selected for quality review over the course of the editing of the sample. The purpose of the quality review is to assess the accuracy of the data, evaluate the work of the editor, and look for improvement opportunities in the editing process. When an editor's return is randomly selected for quality review, a different editor from the same team independently re-edits the return. The two edits of the return are then compared line by line, and discrepancies between the two edits, above a certain tolerance, are stored in the SOI database. For money amount fields, the tolerance is $10; so, money amount fields that differ by $10 or less are not included. However, there is no tolerance for character and code fields. The next step is for a lead editor to review the discrepancies and determine the correct value: the first editor's value, the second editor's value, both, or neither. During the process of reviewing discrepancies, if the first editor value is determined to be incorrect, it is corrected, and the error is charged to the first editor. Then, the reason for the error is determined and coded. There are 32 types of errors; the six most common are shown below.

**Table 1.--Types of Errors**

| Type of Error | Description |
|---|---|
| Affected Entry | Item was incorrect due to an incorrect related item. |
| Improper Allocation | An amount that should have been allocated to another item was not moved or was moved incorrectly. |
| Incorrect Amount | An incorrect amount was entered. |
| Entry on Omitted Form | An item was not edited because the form or schedule was not edited. |
| Omitted Entry | A blank or zero item should have had an entry. |
| Interpretation | Item was edited incorrectly due to being interpreted in a different way than expected. |

Affected entries were the most frequent type of error. These types of error occur when multiple errors are the result of one line item being incorrect. For example, if one line item on Form 1040, such as Salaries, Wages, and Tips, is edited incorrectly, then this causes other line items that use that amount, such as total income, adjusted gross income, and taxable income, to also be incorrect.

**Table 2.--Number of Errors, by Element**

| Element | Number of Errors | Error Rate |
|---|---|---|
| Salaries, Wages, and Tips | 41 | 0.014 |
| Other Income | 51 | 0.018 |
| Total Credits | 13 | 0.004 |
| Income Tax After Credits | 20 | 0.007 |
| Balance Due / Overpayment | 31 | 0.011 |
| Total Depreciation Deduction | 42 | 0.038 |
| Net Investment Income[1] | 19 | 0.023 |
| Tentative Alternative Minimum Tax | 18 | 0.014 |
| Rental Real Estate and Other Passive Activity Net Income/Loss | 21 | 0.027 |
| Other Taxes[2] | 28 | 0.028 |
| Investment Interest[2] | 11 | 0.011 |
| Other Investment Interest[2] | 11 | 0.011 |
| Contract Labor Expense[3] | 24 | 0.021 |
| Utilities Expense[3] | 27 | 0.023 |
| Sole Proprietorship Other Expenses[3] | 109 | 0.093 |
| Net Profit/Loss from Business[3] | 20 | 0.017 |
| Long-Term Gains/Losses from Sale of Capital Assets | 19 | 0.010 |
| Partnership Nonpassive Income | 15 | 0.008 |
| S Corporation Nonpassive Loss | 17 | 0.009 |

[1] Reported on Form 4952
[2] Reported on Schedule A
[3] Reported on Schedule C

## ▶ Study and Limitations

A total of 2,907 returns was selected for regular quality review. Using data from these quality review returns, variables of interest were chosen for this paper. The variables are Salaries, Wages, and Tips; Other Income; Total Credits; Income Tax After Credits; Balance Due/Overpayment; Total Depreciation Deduction; Net Investment Income; Tentative Alternative Minimum Tax; Rental Real Estate and Other Passive Activity Net Income/Loss; Other Taxes; Investment Interest; Other Investment Interest; Contract Labor Expense; Utilities Expense; Sole Proprietorship Other Expenses; Net Profit/Loss from Business; Long-Term Gains/Losses from Sale of Capital Assets; Partnership Nonpassive Income; and S Corporation Nonpassive Loss. These items were chosen by the subject-matter specialists because of the combination of a high number of editor errors and interest in the items.

All returns sampled for the Statistics of Income Individual Tax Return Study are subject to consistency tests. Subject-matter analysts review any returns that fail the consistency tests before the values are considered final. As a result of this review, some values are adjusted; however, there is no information available on these adjustments. The adjusted values replace the original ones.

Several statistics are presented in this discussion of nonsampling error. Net Difference Rate (NDR), t-test, and Index of Inconsistency (IOI) use only the quality review data, while Design Effect (DEFF) uses the entire sample.

## ▶ Simple Response Variance Model

We will consider a simple model that was first proposed by Hansen et al. (1952) and Sukhatme and Seth (1952) for measurement error. Their model specifies that the true value $\mu_i$ (the final value) is different from the observed value $y_i$ (the editor's value) by an unobserved additive error term $\varepsilon_i$. For unit i (i = 1, 2, ... , n), the assumed model is

$$y_i = \mu_i + \varepsilon_i \quad . \qquad (5.1)$$

While we did not measure response error, we adopted these models to our data to measure processing error and estimate bias. The distribution of the editor error variable $\varepsilon_i$ is conceptual; it could be viewed as sampling from a hypothetical population of errors. Thus, the further assumptions for model (5.1) are

$$E[\varepsilon_i | i] = B_i \neq 0$$
$$Var[\varepsilon_i | i] = \sigma_i^2$$
$$E[\varepsilon_i^2] = \sigma^2$$
$$Cov[\varepsilon_i, \varepsilon_j] = 0, \ i \neq j.$$

In words, a systematic bias exists because the mean of the errors is not zero and the error variances are not equal. Also, all errors are uncorrelated. This means that errors made to a return by the first or second editor do not affect other returns edited in the same edit period.

Following Brick et al. (1996), we will assume that the quality review sample is an unrestricted simple random sample, thus

$$E[\mu_i] = \bar{\mu}$$
$$V[\mu_i] = \sigma_\mu^2$$
$$Cov[\mu_i, \mu_j] = 0, \ i \neq j.$$

Under model (5.1), we assume that the first editor's error term no longer averages to zero, possibly due to editor bias, defined as

$$B = \sum_{i=1}^{N} (y_i - \mu_i) \quad . \qquad (5.2)$$

The bias can be estimated by the *Net Difference Rate* (NDR), which is given by

$$NDR = \bar{y} - \bar{\mu} \quad , \qquad (5.3)$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} \mu_i$, and $n$ is the sample size. It can be shown that, if $\mu_i$ is the true value, then the expected value of the NDR is the bias, and its variance exists (Biemer and Atkinson, 1992). Table 3 shows the estimated NDR and t-test values.

**Table 3.--Net Difference Rate and T-Test, by Element**

| Element | NDR | t-test |
|---|---|---|
| Salaries, Wages, and Tips | 5,159 | 0.97 |
| Other Income | -5,895 | 1.11 |
| Total Credits | 3 | 1.73 |
| Income Tax After Credits | -3 | 0.76 |
| Balance Due | 9 | 0.45 |
| Overpayment | -19 | 1.30 |
| Total Depreciation Deduction | -1,016 | 2.43 |
| Net Investment Income[1] | -2,820 | 0.88 |
| Tentative Alternative Minimum Tax | -3,144 | 1.34 |
| Rental Real Estate and Other Passive Activity Net Income/Loss | 1,581 | 1.13 |
| Other Taxes[2] | 186 | 1.41 |
| Investment Interest[2] | -79 | 0.61 |
| Other Investment Interest[2] | 79 | 0.61 |
| Contract Labor Expense[3] | -1,109 | 1.57 |
| Utilities Expense[3] | -43 | 0.15 |
| Profit/Loss from Business Other Expenses[3] | -670 | 0.18 |
| Net Profit/Loss from Business[3] | 842 | 0.59 |
| Long-Term Gains from Sale of Capital Assets | -6,524 | 0.99 |
| Long-Term Losses from Sale of Capital Assets | -5,828 | 2.23 |
| Partnership Nonpassive Income | 461 | 1.68 |
| S Corporation Nonpassive Loss | -512 | 1.82 |

[1] Reported on Form 4952
[2] Reported on Schedule A
[3] Reported on Schedule C

Since the values for the t-test are greater than 1.96 for Total Depreciation Deduction (2.43) and Long-Term Losses from Sale of Capital Assets (2.23), these items have significant bias. This means that the editors are editing these fields differently.

## ▸ Index of Inconsistency

Index of Inconsistency and Design Effect cannot be calculated for those elements with a significant bias because these equations assume the elements have zero bias. For the remaining elements in Table 3 with insignificant bias, we assume the bias is zero, $E[\varepsilon_i|i] = B_i = 0$, and calculate the following statistics:

$$Var[\bar{y}] = Var[\bar{\mu}] + \frac{\sigma^2}{n} = SV + EV. \quad (6.1)$$

The *sampling variance*, SV, is the ordinary variance with no editor error. The *editor variance*, EV, is the variability of returns averaged over conceptual repetitions of editing under the same conditions.

**Table 4.--Index of Inconsistency, by Element**

| Element | IOI |
|---|---|
| Salaries, Wages, and Tips | 0.00184 |
| **Other Income** | **0.18419** |
| Total Credits | 0.00000 |
| Income Tax After Credits | 0.00000 |
| Balance Due | 0.00000 |
| Overpayment | 0.00000 |
| Net Investment Income[1] | 0.00014 |
| Tentative Alternative Minimum Tax | 0.00086 |
| Rental Real Estate and Other Passive Activity Net Income/Loss | 0.00009 |
| Other Taxes[2] | 0.00034 |
| Investment Interest[2] | 0.00002 |
| Other Investment Interest[2] | 0.05339 |
| Contract Labor Expense[3] | 0.00743 |
| Utilities Expense[3] | 0.00870 |
| Profit/Loss from Business Other Expenses[3] | 0.01072 |
| Net Profit/Loss from Business[3] | 0.00476 |
| Long-Term Gains from Sale of Capital Assets | 0.00171 |
| Partnership Nonpassive Income | 0.00005 |
| S Corporation Nonpassive Loss | 0.00007 |

[1] Reported on Form 4952
[2] Reported on Schedule A
[3] Reported on Schedule C

Hansen et al. (1964) define the *Index of Inconsistency* (IOI) as

$$IOI = \frac{EV}{SV + EV}, \quad (6.2)$$

which we use to estimate the proportion of random errors associated with editor error in total variance. The IOI obtains values between 0 and 1.0. Estimated IOI values are shown in the Table 4.

Yu et al. (2000) define that the reliability of the data can be expressed in this equation:

$$r = 1 - IOI. \quad (6.3)$$

In other words, the reliability of an element is the information without the inconsistent portion. All of the elements, except for Other Income, have index of inconsistencies less than .01, which means that they are over 99-percent reliable. Other Income, with the highest Index of Inconsistency (0.18419), is the element with the least amount of reliability, 82-percent, and the largest amount of processing errors.

## ▸ Design Effect

By treating the editors as clusters, the Intra-Editor Correlation Coefficient and Design Effect can be used to measure the editor effect on the variance if the sample was an unrestricted simple random sample.

The Intra-Editor Correlation Coefficient (ρ) measures the correlation between the values that is due to editor error. It is a measure of the similarity of the editors in the way the editors edit a specific element.

Kish (1965) defines the Intra-Editor Correlation Coefficient as

$$\rho_{ed} = \frac{S^2_{between} - \frac{S^2_{within}}{B}}{S^2} \quad (7.1)$$

The ideal range is 0 to 0.1 which indicates no editor variance.

Once the Intra-Editor Correlation Coefficient is calculated, we can use $\rho_{ed}$ to determine the design effect. Design Effect is a measurement of the degree to which an estimate is affected by editor variance,

$$deff = 1 + (B - 1)\rho_{ed} \tag{7.2}$$

where $B$ is the average editor workload or 1,728 returns.

An Editor Design Effect of 1 indicates no increase in variance resulting from the editors. A value of 2 indicates that the variance is doubled.

As Table 5 shows, Overpayment has the largest intra-editor correlation coefficient (0.0124) and design effect (22.40), but one of the smallest Coefficients of Variation. The design effect represents the inflation of variation of the sample if it were treated as a simple random sample with replacement. The design effect for Overpayment can be reduced if editor workload is reduced, but, because the CV is so low, reducing the editor workload in order to reduce the design effect would not be worth the cost.

**Table 5.--Design Effect and Coefficients of Variation, by Element**

| Element | $\rho$ | Design Effect | CV |
|---|---|---|---|
| Salaries, Wages, and Tips | 0.0041 | 8.16 | 0.21% |
| Other Income | 0.0000 | 1.01 | 3.92% |
| Balance Due | 0.0023 | 5.04 | 0.81% |
| **Overpayment** | **0.0124** | **22.40** | **0.38%** |
| Other Taxes[1] | 0.0004 | 1.62 | 4.46% |
| Investment Interest[1] | 0.0005 | 1.94 | 1.73% |
| Long-Term Gains from Sale of Capital Assets | 0.0053 | 10.22 | 1.36% |

[1] Reported on Schedule A

## ▶ Conclusions

This paper was written to estimate the nonsampling error and measure the reliability of the Individual Tax Return Study. Quality Review data were used to measure processing errors and determine how editor error affects the accuracy of specific elements.

From the calculations of Net Difference Rate and Index of Inconsistency, we can conclude that bias can be significantly reduced if we work on the editing procedures for Long-Term Gains/Losses from Sale of Capital Assets, Total Depreciation Deduction, and Other Income. Most of the time, processing errors of several elements can be reduced if the editors concentrate on one element. For example, Other Income has one of the largest Net Difference Rates and the largest Index of Inconsistency, but the smallest Design Effect. In other words, more editors than desired are consistently editing the element incorrectly. Since editors are making similar errors, the data quality can be increased if clearer directions or explanations in the edit manuals are provided. Also, more intense training and examples might lead to smaller processing errors. In addition, this will improve the large positive Net Difference Amount, or overestimate, for Salaries, Wages, and Tips because Other Income allocation is most likely the cause of this problem.

Overall, the editors are producing high-quality work with the exception of specific elements that require more than just transcribing. From the research in this paper, improvement opportunities have become available, and subject-matter analysts can put procedures in place to check the editing quality of specific elements. In addition, editing procedures for elements with high processing errors can be revised and clarified to enhance the accuracy and reliability of the Individual Tax Return Study.

## ▶ References


Biemer, P. and Atkinson, D. (1992), "Estimation of Measurement Bias Using a Model Prediction Approach," *1992 Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 64-73.

Brick, M.; Kim, K.; Nolin, M.J.; and Collins, M. (1996), "Estimation of Response Bias in the NHES: 95 Adult Education Survey," *Working Paper Series*, National Center for Education Statistics, Washington, DC.

Hansen, M.H.; Hurwitz, W.N.; and Madow, W.G. (1952), *Sample Survey Methods and Theory*, John Wiley and Sons, New York, Volume II.

Hansen, M.H.; Hurwitz, W.N.; and Pritzker, L. (1964), "The estimation and interpretation of gross differences and the simple response variance," *Contribution to statistics*, in C.R. Rao (editor) Pergamon Press, Oxford, and Statistical Publishing Society, Calcutta, pp. 111-136.

Kish, Leslie (1965), *Survey Sampling*, John Wiley and Sons, New York, pp. 164-178.

Sukhatme, P.V. and Seth, G.R. (1952), "Nonsampling errors in surveys," *Journal of Indian Society of Agricultural Statistics*, pp. 5-51.

Yu, C.H.; Ohlund, B.; DiGangi, S.; and Jannasch-Pennell, A. (2000), "Estimating the reliability of self-reporting data for Web-based instruction," Arizona State University, Instruction and Research Support.