

Preliminary Results of the National Research Program's Reporting Compliance Study of Tax Year 2001 Individual Returns

Charles Bennett, Internal Revenue Service

For the past 14 years, IRS has relied on the compliance data collected by the Taxpayer Compliance Measurement Program (TCMP) survey of Tax Year (TY) 1988 individual returns to meet a variety of tax administration needs. The data supported national-level measures of reporting accuracy. They were used to update¹ the DIF formulas that IRS used to identify returns for potential audit. New systems, e.g., scoring returns with UI-DIF formulas,² emerged based on analyses of the data.

Had IRS continued to follow the TCMP approach for collecting compliance data, it probably would have surveyed individual returns filed for Tax Years 1991, 1994, 1997, and 2000, and there would be less concern about the age of the data. But many viewed the approach as too burdensome for both taxpayer and IRS. Consequently, while IRS made three attempts to conduct a TCMP survey during the 1990's, it was unable to garner the support of important stakeholder groups, e.g., Congress, and each attempt failed.

In 2000, IRS established the National Research Program (NRP) to develop and implement a new compliance data collection approach—one that (a) minimized (as much as possible) the burden of collecting data without sacrificing data quality, (b) ensured that collected data will meet business objectives and are used as a corporate asset, and (c) involved internal and external stakeholder groups in the design and implementation of reporting compliance studies.

The first NRP study of reporting compliance looked at individual returns filed for TY 2001. A sample of approximately 46,000 returns was randomly selected from a stratified population of returns filed during 2002. Certain information about the taxpayer was associated with each sample return and was then reviewed (also referred to as “classified”) by NRP-trained examiners to identify potential compliance issues. If issues were present, the examiner recorded them and then selected the appropriate audit method—correspondence or face-to-face. All recorded issues had to be questioned by the examiners³ to whom the returns were assigned for audit. The examiners were also expected to identify new issues (including those that benefited the taxpayer) that arose during the course of the audit.

The study was largely completed in the first quarter of Fiscal Year 2005. The following is an overview of the study results to date. Note that IRS is in the early stages of analyzing the data and the results should not be considered final until they have been thoroughly vetted.

As a preface, consider that NRP selected the sample of study returns after IRS's math error processing but before the underreporter programs. Consequently, the study results do not include estimates of the misreporting that is detected by math error processing but do include estimates of the misreporting that would normally be detected by the underreporter or matching programs. The results do not consider any changes that may have been made by Appeals or the courts to examiner-recommended amounts nor do they provide estimates of tax collected. Also, the results were affected by a number of other factors that I describe later in this paper. Finally, these results reflect the "examiner-detected" portion of misreporting. They do not include the nondetected portion of misreporting that would be included in estimates of the gross tax gap.

How many sample returns were selected for each data collection method?

Beginning in August 2002, NRP-trained examiners reviewed case files prepared for each sample return. The purpose of the review was to make an initial assessment of the accuracy of what the taxpayer reported on the return. To the extent possible, the examiner used information⁴ assembled in the case file to substantiate reported items. If all reported items were substantiated and there were no indications of other items that should have been reported, then the examiner "accepted" the return as being accurate. Otherwise, the examiner selected the return for either a correspondence⁵ audit or a field audit requiring an in-person interview with the taxpayer. The results of the review are shown in Table 1.

Initial estimates of the numbers of "accepted" returns and returns selected for correspondence audit were larger than those produced by the study. There were several reasons for this outcome. One was that the final sample design required the selection of more complex, higher-income returns. Another was that classification procedures had not been fully developed when the initial estimates were released. Also, early expectations proved to be overly optimistic concerning the types and numbers of compliance issues that could be handled by correspondence audit.

Table 1.--Distribution of Sample Returns by Data Collection Method

Review [Classification] Result	Number	Percent
Accepted as Accurate	2,535	5.5
Accepted with Adjustment[a]	402	0.9
Previously Audited[b]	186	0.4
Selected for Correspondence Audit	1,817	3.9
Selected for Field Audit	41,067	89.3
Total	46,007	100.0

[a] These sample returns were treated as "accepts" and were not audited. However, small adjustments were indicated by information documents, e.g., W-2 and Form 1099, which did not match the return. The adjustments and their tax impact were captured in the research database.

[b] Certain sample returns were audited prior to being selected by NRP. As a result, the returns could not be reaudited. Rather, the results of the prior audits were collected, reviewed, and added to the database as if they had been collected by an NRP audit.

How accurate is reported total tax liability?

Refer to Table 2. For TY 2001, the estimate of underreported total tax liability (Line 58 of Form 1040) that IRS can detect is \$80.3 billion; the overreported portion is \$4.9 billion. This compares with the estimates for TY 1988 of \$32.3 billion and \$2.1 billion respectively.

Table 2.--Estimates of Change to Total Tax Liability

Tax Year	Underreported	Overreported
1988	\$32.3B	\$2.1B
2001	\$80.3B	\$4.9B

The resulting voluntary reporting rate⁶ (VRR) is 92.4 percent, a rate somewhat lower than the 93.2 percent estimated for TY 1988 returns which suggests that a real but modest deterioration in voluntary compliance has occurred since 1988. Table 3 provides the rates determined by the last five compliance studies.

Table 3.--Voluntary Reporting Rates – Individual Income Tax Returns

Tax Year	VRR
1979	91.6%
1982	92.7%
1985	92.0%
1988	93.2%
2001 (NRP)	92.4%

Given the new approach used by NRP for collecting compliance data, is VRR based on NRP comparable to the measure produced by TCMP?

The design of the NRP study included a separate small sample of TY 2001 individual returns to be audited in much the same manner as would have been done under TCMP. The purpose for this design requirement was to provide data that would allow IRS to observe if there is a statistically significant difference in the VRR produced by NRP and TCMP. The size of the “calibration” sample was determined to provide an estimate of VRR that, if different from the NRP VRR by more than 2 percentage points, would indicate a difference that was statistically significant and could be attributed to the change in data collection approach.

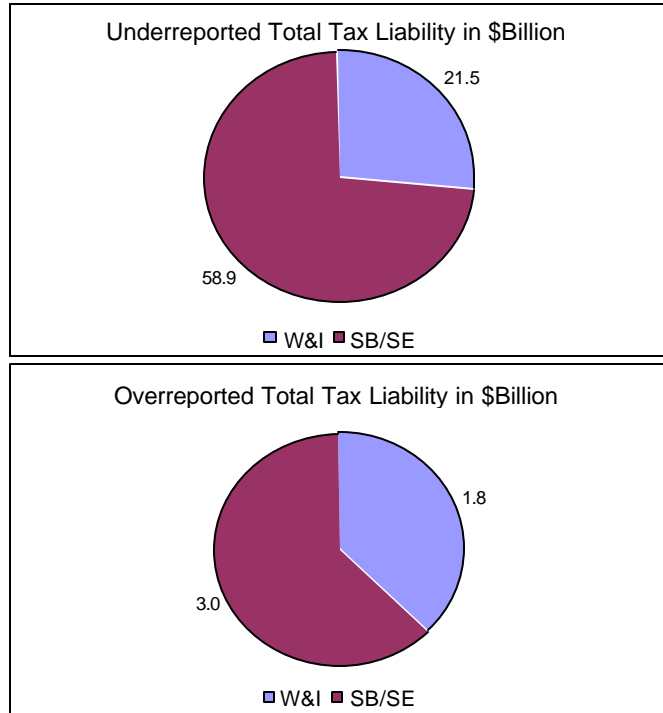
The calibration sample was divided somewhat equally among returns that, initially, were classified as accepted, selected for correspondence audit, or selected for face-to-face audit. Except for a few cases,⁷ all calibration sample returns were then audited using the face-to-face technique, and potential issues were determined without considering any dollar tolerances.

To date, the calibration sample results appear to indicate slightly more than two percentage points’ difference in VRR. What is more, the results suggest that NRP detects more net noncompliance than TCMP. This is counterintuitive. More analysis is needed to determine if the calibration results were adversely affected by the small size of the sample, incorrect weighting of the sample, or other conditions, e.g., data anomalies discussed later in this paper. Consequently, until we have a better explanation for the calibration sample results, we will assume that the VRR is statistically the same as that produced by the larger noncalibration sample.

What is the distribution of tax change by the IRS Operating Division?

The estimates of underreported and overreported total tax liability that IRS can detect were also tabulated by the IRS Operating Division—SB/SE and W&I—that had administrative responsibility for the return. Figure 1 reveals that SB/SE returns account for \$58.9 billion of the underreported total tax; W&I returns, \$21.5 billion. The overreported portion was \$3.0 billion for SB/SE returns; \$1.8 billion for W&I returns.

Figure 1.--Tax Change Distribution by IRS Operating Division



What are the VRR measurements by the IRS Operating Division?

Refer to Table 4. While the overall VRR for TY2001 returns is 92.4 percent, that rate is not the same for both SB/SE returns and W&I returns. SB/SE returns had a VRR of 90.4 percent indicating a lower level of reporting accuracy than the 95.2 percent observed for W&I returns.

Table 4.--Voluntary Reporting Rates – TY2001 Individual Income Tax Returns

Tax Year	VRR
Wage and Investment	95.20%
Small Business/Self-Employed	90.40%

How accurate is the amount reported as overpaid or owed?

In the past, IRS used a variety of measures to describe the accuracy of information reported by taxpayers on their tax returns. The measures included, among others, VRR and VCL (voluntary compliance level), VRP (voluntary reporting percentage) and NMP (net misreporting percentage), and VCR (voluntary compliance rate) and NCR (noncompliance rate).

Most often, IRS used VCL and VRR to measure overall return accuracy (to the extent that examiners could determine it). Of the two, VRR is now preferred because it does not disregard overreporting of tax liability as VCL does. Also, the acronym "VCL" may imply that the measure accounts for more than reporting accuracy, i.e., payment and filing compliance. It does not.

VRP and NMP are both measures of reporting accuracy for a particular return line item. VCR and NCR are measures associated with tax gap estimates.

Within this framework of measures, VRR measures the accuracy of total tax but does not consider the accuracy of reported payment amounts, including refundable credits, e.g., earned income credit and additional child tax credit. Consequently, a measure is needed to quantify the accuracy of the net tax liability (total tax less payments) reported by the taxpayer.

Table 5 provides a measure of the accuracy of reported net tax liability by categorizing returns as "underreported," "overreported," or "accurate." More than 52 percent of returns have correctly reported net tax liabilities. The table also shows the effect of redefining what might be considered an "accurate" return to include those with a small change, such as \$100 or less, to the net tax liability. Based on that definition, more than 63 percent of returns have correctly reported net tax liabilities.

Table 5.--Distribution of Individual Income Tax Filers into "Underreporter," "Overreporter," and "Accurate" Categories, Tax Year 2001

Compliance Category	Percent of the Population	
	"Accurate" Defined as No Change in Net Tax Liability	"Accurate" Defined as Change in Net Tax Liability <= \$100
Underreported net tax liability	39.8%	32.1%
Overreported net tax liability	8.0%	4.5%
Accurately reported net tax liability	52.2%	63.4%
Total	100.0%	100.0%

What were the most frequently misreported line items for 2001?

Table 6 ranks the weighted number of required corrections (reporting error) to return line⁸ items. At just over 17 million, there were more required corrections to cash contributions reported on Schedule A, line 15 than any other single item of income, deduction, tax, or credit. By comparison, interest income was the most corrected item on 1988 returns with 15.5 million corrections, and cash contributions was second with 11.1 million required corrections. Note that the calculation of error rate does not take into account the size or direction of the error. For example, a return with an underreported EITC of \$1 would be treated the same as a return with a fraudulent EITC overclaim of \$2,000.

Table 6.--Top 10 Misreported Line Items for TY2001

Form/Schedule	Line Number	Description	Estimated # of Returns With Reporting Error on Line	Total # of Returns With Entry[c] on Line	Error Rate %
SCH A	15	Cash Contributions	17,049,517	37,855,184	45.0
1040	8	Interest Income	13,639,410	67,479,816	20.2
1040	6	Exemptions	9,670,939	47,660,219	20.3
SCH A	6	Real Estate Taxes	8,800,248	38,716,754	22.7
1040	61	Earned Income Tax Credit	8,413,407	19,593,121	42.9[d]
1040	7	Wages and Salaries	8,355,798	111,227,450	7.5
SCH A	16	Noncash Contributions	8,350,840	22,585,276	37.0
SCH A	7	Personal Property Taxes	6,907,869	19,953,863	34.6
SCH C	1	Gross Receipts	6,551,827	17,543,163	37.3
SCH C	27	Other Expenses	6,348,485	10,731,171	59.2

[c] Source: IRS Statistics of Income

[d] The estimated number of returns with a reporting error for EITC and the resulting error rate are not directly comparable to published estimates that were based on the 1999 EITC Compliance Study because data collection and analysis methods were different. For example, the 1999 EITC study did not add together under- and overreported EITC and did not measure underreported EITC among taxpayers who did not claim the credit. Also, the 1999 EITC study reported upper- and lower-bound estimates to reflect different assumptions about the accuracy of claims filed by taxpayers who failed to respond to a request for examination.

What were the top 10 misreported line items by amount?

Table 7 ranks the weighted net amounts of corrections (misreported amounts) by line item. The net correction to Schedule C gross receipts ranks as the number-one misreported item with \$48.6 billion underreported. For 1988, this item was also number one with \$20.6 billion underreported. Care should be taken, however, in comparing these estimates with those based on other studies. For example, the \$13.1 billion net correction to EITC is not comparable to the “unrecovered overclaim” reported for the EITC for Tax Year 1999 because the NRP estimate does not account for the effect of EITC examinations. In addition, the 1999 EITC study reported upper- and lower-bound estimates to reflect different assumptions about the accuracy of claims filed by taxpayers who failed to respond to a request for examination. Finally, the NRP estimates often reflect corrections that may be considered questionable. See the following discussion on data anomalies.

Table 7.--Top 10 Misreported Line Items by Amount

Form/Schedule	Line Number	Description	Amount Misreported (\$ billions)
SCH C	1	Gross Receipts	48.6
1040	21	Other Income	34.1
SCH D	1	Short-term Capital Gains and Losses	29.5
1040[e]	7	Wages and Salaries	15.8
SCH A	15	Cash Contributions	13.1
1040[f]	61	Earned Income Credit	13.1
SCH C	27	Other Expenses	13.1
SCH D	8	Long-term Capital Gains and Losses	13.1
SCH E	28bi	Partnership Nonpassive Loss	10.1
SCH A	1	Medical Deduction	8.9

[e] Includes Form 1040A, line 7 and Form 1040EZ, line 1.

[f] Includes Form 1040A, line 39 and Form 1040EZ, line 9.

Why do anomalies exist in the NRP data?

The NRP study of reporting compliance was designed to meet the varied needs of many customers, some known and some anticipated, for data that can be used to measure the accuracy of reported income, deductions, taxes, credits, and payments. For certain customers, it was sufficient to know only the tax change associated with overall misreporting. For others, it was important to collect data that described return accuracy at the line-item level. While the study accomplished both objectives, NRP did not foresee the effect that certain design factors would have on the data, creating what we call data “anomalies.”

One of the factors that contributed to the presence of data anomalies was the examiner use of IRS's Report Generation System (RGS). Prior to NRP, examiners had used this application to prepare their audit reports, and the reports required the capture of both reported and corrected return data. So, NRP viewed RGS as an opportune, less burdensome way to collect audit results for study cases without having examiners complete both an audit report and a study checksheet.⁹ NRP built a study infrastructure that passively collected the data from the RGS case records and loaded the data into a research database located at IRS's Detroit Computing Center. Unfortunately, NRP did not have full knowledge of RGS's internal processes and data stores and ultimately had to concede that some of the expected data would be missing and some of the collected data would be difficult to understand.

What data anomalies are known and what is being done to control them?

The first data anomaly concerns study cases that had a change in filing status from nonjoint, i.e., single, married-filing-separate, or head of household, to a filing status of married-filing-jointly. For the NRP study to date, there are 135 cases that meet this condition. Under TCMP, these types of cases were excluded from study results due to the difficulty in weighting the observation when both the sampled and nonsampled spouse filed a return. The obvious problem is how to deal with the income, deductions, etc., for the non-NRP spouse captured as corrections to the NRP-selected return. NRP is currently collecting data on returns as originally filed for the non-NRP spouses, as well as the audit results for those returns, but has not yet incorporated the results into the database. One of the interesting facts uncovered by this effort is that 70 of the 135 cases involved spouses who did not file.

Another anomaly involved IRS processing of amended TY 2001 returns filed by NRP-selected taxpayers. In prior compliance studies, IRS generally did not process amended returns filed when the original return was being audited. Instead, the amended return was routed to the examiner who considered the information reported on the amended return in determining the final set of corrections made to the original return. During the NRP study, amended returns were received and actually processed by IRS after the audit of the original return was begun. This means that the examiner had to consider the “corrected” amounts on the amended return as the starting point for the audit rather than what was reported on the original return. It is possible to identify which NRP returns were amended and determine if the 1040X was filed after IRS contact. With this information, the “corrections” based on the amended returns can be added to the adjustments made by the examiner to determine the accuracy of the original return.

How examiners handled misclassified items could also result in a data anomaly. Instructions were given to examiners to avoid reclassifying amounts; but, if necessary (a condition often required by RGS to calculate tax liability), examiners could make such a correction as long as it was properly described with a specific reason code. For example, if the taxpayer reported self-employment income as wage income, the examiner had to make at least two adjustments: one to reduce wage income, the other to increase Schedule C gross receipts. RGS required both adjustments to calculate SE tax properly. For this situation, the examiner was instructed to select reason code “14” or “10” depending on whether the reporting was intentional. Unfortunately, examiners did not always select the correct reason code or selected it for one issue, e.g., wages, but not the other, Schedule C gross receipts. Consequently, controlling for adjustment amounts that are based on misclassified items will be problematic.

Finally, examiners frequently came across compliance issues requiring “multi-part” corrections. An example is one in which an examiner detects both gambling winnings and gambling losses for a taxpayer who reported neither. The net effect of these adjustments is often a minimal change in tax liability but a significant change in the accuracy of Form 1040, line 21, “other income,” and Schedule A, line 27, “other miscellaneous deductions.” Considered separately, the accuracy of each line item could raise concerns, considered together, maybe not as much concern. Another example of a “multipart” correction is one in which the examiner detects unreported sales of inventory. Rather than making one adjustment to include the net profit from the sale of inventory, the examiner would typically make one adjustment to increase sales receipts and at least one other adjustment to increase the costs of goods sold.

Again, looking at individual line item results may not always give a clear picture of reporting accuracy.

Has NRP made a difference in how IRS operates?

While it is premature to expect changes in how IRS operates that are based on information obtained from the individual return study, NRP has an impact on processes used in IRS's audit program. NRP pioneered two processes in particular. "Case building" was an NRP design element that has since become a standard procedure used by IRS to prepare returns for classification and audit. In addition, NRP required that examiners not only use RGS, but that they do so in a LAN environment for audits of NRP sample returns. This requirement virtually eliminated the need for transmitting audit results by diskette and provided an effective way for storing results in (and retrieving results from) a central location. During the individual study, the SB/SE Operating Division expanded this requirement to include all audits, not just those involving NRP returns.

Has IRS used the results of the individual filer study?

To date, the results of the NRP individual filers study have been used to make preliminary estimates of the gross tax gap. Also, IRS has used the results to update the DIF formulas that score individual tax returns for audit potential. The new formulas will score returns filed by taxpayers beginning in January 2006. Other analyses of the NRP data are under way.

Endnotes

- ¹ Updated formulas were used by IRS to score returns filed and processed beginning in January 1993.
- ² The formulas produce a score of the probability that a return will have unreported income.
- ³ Examiners were not allowed to audit the returns that they "classified."
- ⁴ Case files included data from tax returns for TY 1998, 1999, and 2000; information returns (detailed records for TY 2001 and summary records for 2 prior years), certain Master File transactions, Dependent Database, Currency Banking and Retrieval System, Choicepoint (compilation of publicly available data), and Net Basis (product of Net Worth, Inc. that provides valuations for cost/basis of securities).

- ⁵ To qualify for a correspondence audit, only certain types of compliance issues could be present, and there was a limit on the number of issues. If the limit was exceeded, then the case was selected for field audit.
- ⁶ VRR is the percentage of the total tax required to be reported that taxpayers voluntarily reported on their timely-filed returns. $VRR = \frac{\text{reported total tax}}{\text{sum of reported total tax and net change to total tax}}$.
- ⁷ Returns that were previously accepted by classification and filed by taxpayers that examiners could not locate were treated as no-change audits.
- ⁸ Certain lines, e.g., $\frac{1}{2}$ SE tax deduction, total itemized deductions, SE tax, were frequently corrected but were not included in the table because their corrections were the result of other lines being corrected.
- ⁹ Data collection method employed by TCMP.