2007 The IRS Research Bulletin

Publication 1500

The IRS Research Bulletin

Proceedings of the 2007 IRS Research Conference

Research
Analysis
Statistics

Impact of Taxpayer Representation on the Outcome of Earned Income Credit Audits

Jeff A. Wilson, Tom Beers, Amy Ibbotson, Mike Nestor, Mark Hutchens, Carol Hatch, and Mark Everett, Taxpayer Advocate Service, IRS Office of Research and Analysis

The IRS administers the Earned Income Credit (EIC) to millions of taxpayers each year.[1] An important aspect of effective tax administration is to ensure the accuracy of the EIC claims. One way the IRS does this is by auditing some of the returns filed. The outcome of the audit presumably validates the taxpayers' eligibility for the EIC. As a matter of fairness and effective tax administration, the IRS must work with taxpayers, and their representatives, to ensure that the EIC is accurately claimed.

The Taxpayer Advocate Service (TAS) wanted to look at these EIC audits to determine if using a representative impacted the outcome of the audit.[2] Specifically, TAS wanted to know if the use of a representative enabled a taxpayer to keep his or her EIC, or at least retain a larger amount of it after the audit.

EIC audits represent approximately 43 percent of all IRS individual taxpayer audits.[3] The vast majority of these taxpayers do not have professional representation during the audit. This is perhaps not too surprising, given the income level of these taxpayers and their likely unfamiliarity in dealing with the IRS on issues involving complicated matters of tax law. Anecdotal reviews of EIC audits where EIC was disallowed show that, frequently, there is no significant evidence that the taxpayer was ineligible. Instead, the taxpayer failed to prove EIC eligibility.[4] For example, when asked to provide school records to verify the 6 months residency requirement, taxpayers often submit records for a single school year. Given that a typical school year overlaps 2 calendar years, this information is insufficient to prove residency to the IRS, but it is not evidence that the taxpayer is ineligible for the credit.

The law clearly places the burden of such proof on the taxpayer, but, if the taxpayer cannot sufficiently understand the rules, reaching the goal of a

[1] Over 21 million filers claimed EIC in Tax Year 2002, EIC Fact Sheet for TY02 as of 12/31/03.

[2] For example, attorney, certified public accountant, enrolled agent, unenrolled agent, etc.

[3] IRS FY 2005 Data Book Table 10 (42.9 percent).

[4] W&I Research review of 43 closed EIC audits.

correct audit outcome is brought into question. TAS is compelled to ask if the lack of representation during an audit puts these taxpayers at an inherent disadvantage over those taxpayers who are represented.

Ideally, the IRS would be able to reach the right outcome on EIC eligibility regardless of the presence of a representative.[5] As we will see in the body of this report, the use of a representative does appear to have a significant impact on the outcome of the audit and the amount of the EIC retained by the taxpayer. This finding suggests the IRS must work harder, and smarter, to reach an accurate resolution of EIC eligibility issues, particularly when the taxpayer does not have a representative.[6]

Given scarce resources, the IRS and taxpayers will be challenged to find a way to better verify EIC eligibility in an audit environment. The IRS simply cannot provide a representative for each taxpayer. TAS believes this study compels the IRS to find new ways of reaching out to those taxpayers who do not have representation. The awarding of EIC to ineligible taxpayers costs the Government billions of dollars; however, disallowing EIC to those taxpayers truly eligible for the credit is negatively impacting their already fragile financial well-being.

Background

Prior IRS studies indicate that a significant proportion of claimants historically have not been entitled to the EIC. For example, of the estimated $31.3 billion in EIC claims made by taxpayers who filed Tax Year 1999 returns in 2000, it is estimated that between $8.5 billion and $9.9 billion (27.0 percent to 31.7 percent) should not have been paid.[7] These estimates were derived by auditing a sample of 3,457 taxpayer returns that claimed the EIC. TAS is interested in knowing if taxpayers would have fared better (i.e., kept their EIC or lost less of their EIC) if they had obtained representation.

TAS recognizes the critical role that auditing serves in tax administration. IRS audits help ensure taxpayer compliance and protect the tax revenue base. However, TAS is concerned by the findings from recent focus groups and targeted interviews with taxpayer representatives regarding barriers taxpayers face during IRS audits. TAS conducted focus groups with taxpayer representatives at the 2005 IRS tax forums and also initiated targeted interviews with Low Income Taxpayer Clinic (LITC) attorneys to discuss

[5] After controlling for self-selection by taxpayers who use a representative, the IRS would presumably find similar rates of EIC eligibility.

[6] The IRS must also ensure that represented taxpayers do not unfairly receive EIC.

[7] Compliance Estimates for Earned Income Tax Credit Claimed on 1999 Returns, Department of the Treasury, Internal Revenue Service, February 28, 2002. Tax Returns were filed in 2000 for TY1999.

problems with audit processes relevant to EIC. In particular, these groups sought to learn what barriers the representatives foresaw that prevented IRS and taxpayers from reaching the correct outcomes on EIC eligibility and amounts claimed. The representatives identified several barriers including inconsistent IRS requests for documentation; lost paperwork; and poor communication.[8]

TAS is concerned that these barriers are preventing the IRS from treating taxpayers fairly. In particular, TAS wants to achieve a proper balance between EIC compliance and accurately determining taxpayers' eligibility for the EIC. If these barriers are preventing the IRS and taxpayers from accurately determining the correct amount of EIC, the IRS may be inadvertently denying taxpayers a credit they are legitimately entitled to receive.

EIC filers have several attributes that may hinder their ability to respond effectively to an audit.[9] These attributes may impede communication and understanding of requests made by the IRS during an audit of the taxpayer's EIC. These problems are exacerbated by barriers raised in the aforementioned focus groups and interviews. TAS is concerned that these taxpayer attributes, and the previously discussed barriers, are leading the IRS to improperly deny taxpayers their EIC.

TAS recognizes the repercussions this may have on tax administration. The National Taxpayer Advocate tasked TAS Research to explore the following issue:

> *Do taxpayers who have representation fare better in EIC audits than those who do not have representation?*

An affirmative answer to this question would highlight the need to reformulate IRS compliance programs that verify the EIC in such a way as to minimize the use of audits and/or modify the way the audits are conducted. This has the potential to impact tens of thousands of taxpayers who claim the EIC.

---

[8] The National Taxpayer Advocate's Findings from Correspondence Examination Focus Groups, IRS Tax Forums June-September 2005, December 2005 and Taxpayer Advocate Service's Challenges for Taxpayers Claiming the Earned Income Tax Credit (EITC), From Interviews with Low Income Tax Clinics, September 2005.

[9] Attributes of EIC filer include: less likely to speak English, less education, and lower income levels. See "Playing by the Rules, but Losing the Game–America's Working Poor," Urban Institute http://www.urban.org/publications/410404.html (last viewed May 31, 2007). These attributes suggest that EIC taxpayers may be less likely to understand IRS correspondence and less able to afford representation (i.e., power of attorney) with IRS.

Research Methods

The population studied in this analysis was comprised of TY 2002 returns audited for EIC issues. EIC returns were selected for audit through various means, including Dependent Database (DDb) processing, Discriminate Income Function (DIF), and EIC Recertification procedures.[10] Tax Year 2002 was chosen because it includes the effects of significant EIC tax law changes implemented that year. Additionally, using TY2002 allowed sufficient time to review case activities that occurred subsequent to the close of the initial audit. For this study, cases were selected from the Automated Information Management System (AIMS) closed case database by project code. The list of project codes used to determine EIC audit cases was obtained from the EIC Program Office. Some returns in these project codes were determined to have never claimed or received EIC, according to IRS Masterfile data, and were thus removed from our analysis. Some additional returns were removed from our study because insufficient data were available for analysis.[11]

The AIMS population data were supplemented with other individual tax return data to obtain such items as amount of EIC claimed by the taxpayer and allowed by the IRS, as well as income information from the return and such entity items as filing status and return preparation method. Gender data were also obtained from the IRS information received from the Social Security Administration. Lastly, the Compliance Research Initiative Tracking System (CRITS) was utilized to obtain additional data necessary to analyze the outcome of the audits.[12] Most notably, the IRS Masterfile transactions for the credit and debit of EIC, and tax were obtained.[13] These transaction data were utilized to determine the amount of EIC claimed by the taxpayer, allowed by the IRS during return processing, and ultimately allowed after the initial audit of the return. Masterfile transaction code data were also utilized to verify the presence of representation during the audit.

The population data were then split into two groups, those taxpayers with representation during the audit and those without representation. The determination of whether a taxpayer was represented during audit was made from data obtained from a special extract of the Centralized Authorization File (CAF) for TY 2002. The CAF data also identify the type of representation.[14] The CAF data were also cross-referenced with the Masterfile transac-

[10] If EIC for any year after 1996 was denied or reduced for any reason other than a mathematical or clerical error, a Form 8862 is required to be filed with the next tax return if claiming EIC with qualifying children.

[11] See data limitations in the following section.

[12] CRITS data contain current IRTF and Masterfile data elements.

[13] The IRS posts debits (i.e., tax assessments) and credits (e.g., EIC credit) to a taxpayer's account with different codes so that the type of each debit or credit may be clearly identified. Separate codes are also used to denote other account activity, such as the authorization of a representative for a taxpayer.

[14] For example, attorney, certified public accountant, enrolled agent, unenrolled agent, etc.

tion codes indicating the presence or removal of a representative. In a few cases, the CAF and Masterfile data were discrepant, and these cases were removed from the study population.[15]

The Examination start date and disposition date from the AIMS data were used to split transaction code data from the Masterfile into four time periods—before audit, first audit, second audit, and after audit.[16, 17] Transaction codes with cycles posting before the Examination start date were included in the before audit time period. Transaction codes with cycles posting between the Examination start date and the first audit disposition date were included in the first audit period. The second audit period includes transaction codes posting between the first audit disposition date and the second audit disposition date. Transaction codes after the last audit disposition date were incorporated in the after audit period.

Representation, EIC change, and tax change were defined for each of the time periods.[18, 19, 20] This report includes analysis using the before audit and first audit periods. TAS Research may, in the future, analyze the second audit and after audit periods. TAS Research also plans to look at the outcome of appealed EIC audits for represented and unrepresented taxpayers.

Unless otherwise noted, the findings are based on a dataset containing 328,429 taxpayers. Of this number of returns, only 11,411 (3.6 percent) were represented in the original audit. The original study data contained over 360,000 returns with an EIC project code.[21] However, as described in the following limitations section, several circumstances necessitated the removal of returns from the study.

Limitations

When analyzing the data, TAS Research discovered several anomalies in the data for the population of TY 2002 EIC taxpayers who were audited. Based

[15] 68 cases were removed for this reason.

[16] The Examination start date and disposition dates were converted to cycle posting dates using Document 6209, Integrated Data Retrieval System (IDRS) and Automated Data Processing (ADP) Book. The Examination start date was in an YYYYMM format. To convert to a posting cycle, we assumed the audit started on the first day of the month. In addition, four cycles were added to both the start and disposition dates to account for the time delay between AIMS and IRTF postings. TAS Research based this decision on analysis of the data and consultation with knowledgeable Examination/AIMS personnel.

[17] There were 545 taxpayers with two audit indicators, indicating the return was reviewed twice by the IRS.

[18] Representation is noted on a tax module by transaction codes 960, 961, and 962.

[19] EIC change was determined from transaction codes 764, 765, and 768.

[20] Transaction codes 290, 291, 294, 295, 298, 299, 300, 301, 304, 305, 308, and 309 were used to compute tax change.

[21] Data extracted as of June 2006.

on this analysis, returns with the following characteristics were eliminated from the population:

1. Duplicate AIMS taxpayer records;
2. Taxpayers who did not claim EIC on their tax returns or did not check the box on their tax returns to have the IRS compute EIC for them;
3. Taxpayers with a nonexamined (survey, accepted as filed) disposal code;
4. Taxpayers with undelivered audit notices were removed since they never participated in the audit process;[22]
5. Taxpayers whose filing status was Married Filing Separate (MFS). This group was eliminated due to incomplete information on changes to filing status (i.e., MFS to Head of Household);
6. Taxpayers with missing tax return data on the CRITS or Compliance Data Warehouse (CDW);
7. Taxpayers who filed a Form 1040X were removed due to insufficient data regarding the claim; and
8. Taxpayers with inconsistent representation information on Masterfile (transaction code 960) and CAF data.[23]

During data analysis, TAS Research also observed instances in the data where the taxpayer did not have qualifying children, but EIC before audit and the change in EIC due to the audit were greater than the maximum amount of EIC allowed for taxpayers without children. To correct for this anomaly, we updated the number of qualifying children based on the EIC Table in Publication 596 for TY 2002.[24]

[22] Undelivered mail was determined from the AIMS technique code field.

[23] The CAF data file contains information from Form 2848, *Power of Attorney and Declaration of Representative*. For purposes of this study, a taxpayer was considered represented if the representative authorization appeared on both the CAF and the IRS Masterfile.

[24] There were 22 of these cases.

Objectives

1. Determine if taxpayers with representation in EIC audits are more likely to be determined eligible for EIC (and to have a higher no-change rate) than taxpayers without representation in EIC audits.[25]
2. Determine if taxpayers with representation in EIC audits retain a greater proportion (measuring the proportion retained will help guard against the bias of one group claiming more EIC than the other) of the EIC originally claimed than taxpayers without representation in EIC audits.
3. Determine if the tax recommended (this measure will allow for an analysis of the effect of representation on related issues (i.e., filing status) also examined during EIC audits) for taxpayers with representation in EIC audits is less than the tax recommended for taxpayers without representation in EIC audits.
4. Determine the extent of the effect (measured by a regression analysis of EIC dollars reduced by audit) of representation on the outcome of EIC audits.
5. Compare return and other demographic characteristics of the EIC audit population with representation to those without representation.

Research Findings

Objective 1: Determine if taxpayers with representation in EIC audits are more likely to be determined eligible for EIC (and to have a higher no-change rate) than taxpayers without representation in EIC audits.

Represented taxpayers are twice as likely to be found eligible for EIC and to have no changes made to their EIC.

[25] No-changes audits are those in which the IRS does not adjust the returns based on the audit findings.

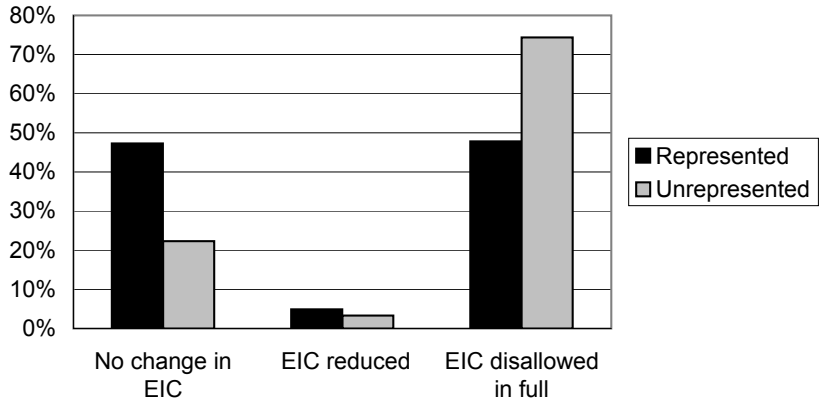
Table 1 depicts the percentage of taxpayers who retained at least some EIC after audit. Clearly, represented taxpayers were much more likely to retain their EIC after audit than unrepresented taxpayers. In fact, taxpayers who used a representative during the audit process are more than twice as likely to be determined EIC-eligible when compared to taxpayers without representation.[26]

Table 1. EIC Retained / Disallowed During Audit[27]

Percentage of Taxpayers with:	Represented	Unrepresented
No change in EIC[28]	47.3%	22.3%
EIC reduced	4.9%	3.3%
EIC disallowed in full	47.8%	74.4%
	100.0%	100.0%

Source: IRTF TY2002 and CAF for TY2002

Figure 1. Impact of Representation on EIC Allowed During Audit



Source: IRTF TY2002 and CAF for TY2002

Likewise, Table 1 depicts that the "no change" rate for represented taxpayers is also more than double that for unrepresented taxpayers (47.3

[26] The percentage of taxpayers retaining some EIC is 52.2 percent for taxpayers with representation compared with 25.6 percent retaining some EIC for taxpayers without representation. The difference between represented and unrepresented taxpayers is statistically significant at level .0001 (one-sided t-test).

[27] The difference between represented and unrepresented taxpayers is statistically significant at level .0001 (one-sided t-test).

[28] This "no change" rate includes taxpayers who actually received additional EIC as a result of the audit. This includes 1.0 percent overall of represented taxpayers and 0.3 percent overall of unrepresented taxpayers.

percent versus 22.3 percent). A comparison of the data in this table indicates that relatively few taxpayers remain eligible for EIC but receive a reduced amount. This circumstance is likely attributable to the fact that EIC eligibility is mostly based on hard and fast rules regarding a child's relation to and residency with a taxpayer. Accordingly, little middle ground remains for a partial allowance of EIC, underscoring the importance of the IRS reaching a correct audit determination.

The type of representative also has an impact on the change in the EIC received, as shown in Table 2 below. More than half, 52 percent, of taxpayers represented by attorneys and CPAs retain the full amount of their EIC claims. Taxpayers represented by generally less sophisticated, unenrolled agents retained EIC for their clients only 44.5 percent of the time. This finding implies that representatives with more training are better able to successfully represent their clients and suggests that minimum standards should be considered to enable a representative to practice before the IRS. Nevertheless, it should be noted that taxpayers using representatives with fewer credentials still achieve considerably more favorable results than taxpayers without representation.

Table 2. EIC Retained/Disallowed During Audit, by Type of Representative

	Attorney or CPA	Enrolled Agent	Unenrolled Agent	Other Representative[29]	Total
Count	3,617	2,300	4,228	1,266	11,411
Percentage of taxpayers with no change in EIC[30]	52.0%	48.5%	44.5%	40.8%	47.3%
Percentage of taxpayers whose EIC was reduced	4.4%	4.6%	5.6%	4.7%	4.9%
Percentage of taxpayers whose EIC was disallowed in full	43.6%	46.9%	50.0%	54.6%	47.8%
	100.0%	100.0%	100.1%*	100.1%*	100.0%

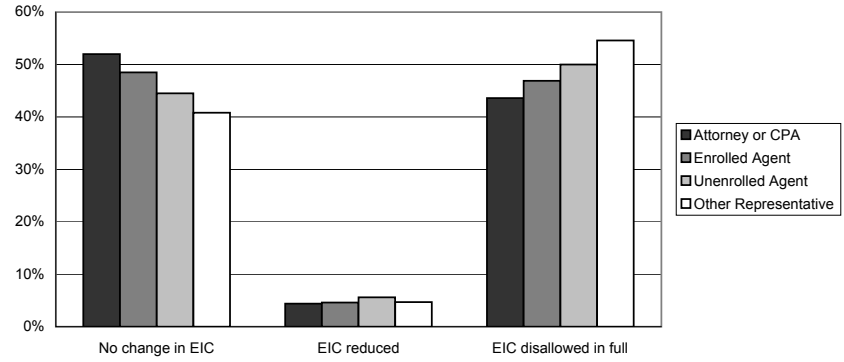
* Numbers add to more than 100.0 percent due to rounding.

Source: IRTF TY2002 and CAF for TY2002.

[29] The "Other Representative" category includes full-time employees (officers) of the taxpayer's organization, family members, and enrolled actuaries.

[30] This "no change" rate includes taxpayers who actually received additional EIC as a result of the audit. This includes 1.0 percent overall of represented taxpayers and 0.3 percent overall of unrepresented taxpayers.

Figure 2. EIC Retained/Disallowed During Audit, by Type of Representative



Source: IRTF TY2002 and CAF for TY2002.

Objective 2: Determine if taxpayers with representation in EIC audits retain a greater proportion of the EIC originally claimed than taxpayers without representation in EIC audits.

Represented taxpayers retain more of their EIC.

The prior section focused on the percentage of taxpayers whose EIC was reduced or remained the same. Another way to analyze the data is to look at the percentage of EIC dollars retained. Table 3 below shows that taxpayers with representation retained, on average, 50.2 percent of their EIC versus 24.0 percent for taxpayers without representation.

Table 3. Portion of EIC Retained During Audit[31]

	Represented	Unrepresented
Average percentage of original EIC retained	50.2%	24.0%

Source: IRTF TY2002 and CAF for TY2002

As in the prior section, taxpayers using representatives with more credentials receive more favorable outcomes. Table 4 shows that taxpayers who used an attorney or CPA retained 54.8 percent of their EIC during the

[31] When looking at taxpayers whose EIC was not changed, the difference between represented and unrepresented was still more than 2 to 1 (4.9 percent represented to 2.0 percent unrepresented). The difference between represented and unrepresented taxpayers is statistically significant at level .0001 (one-sided t-test).

audit, over 7 points higher than unenrolled agents and over 10 percentage points higher than for other representatives.

Table 4. Portion of EIC Retained During Audit, by Type of Representative

	Attorney or CPA	Enrolled Agent	Unenrolled Agent	Other Representative	Total
Count	3,617	2,300	4,228	1,266	11,411
Average percentage of original EIC retained	54.8%	51.1%	47.6%	43.5%	50.2%

Source: IRTF TY2002 and CAF for TY2002

The number of qualifying children is one of the key determinants of the amount of EIC to which a taxpayer is entitled. Table 5 shows that represented taxpayers retain a greater share of their EIC over unrepresented taxpayers regardless of the number of children.

Table 5. Portion of EIC Retained During Audit, by Qualifying Children

	Represented	Unrepresented
No Qualifying Children	100.8%[32]	56.7%
One Qualifying Child	47.3%	19.6%
Two Qualifying Children	52.0%	28.1%

Source: IRTF TY2002 and CAF for TY2002.

Another consideration is the impact a given change has on the taxpayer. For example, perhaps the 50.2 percent of EIC retained by represented taxpayers is offset by the absolute amount of the reviewed credit because unrepresented taxpayers have higher claimed amounts. (In other words, disallowing in full an EIC of $100 will likely have less effect on a taxpayer than reducing a $4,000 credit by half).

Table 6 shows the averages for EIC reviewed, changes, and net final amount. The first observation we can make is that represented and unrepresented taxpayers have similar before-audit EIC amounts, a difference of only $66. Second, the average EIC disallowed is $669 higher for unrepresented taxpayers. The overall result is that taxpayers with representation retain $735 more than taxpayers without representation, despite having an initial EIC of only $66 more.

[32] These taxpayers received more EIC after audit than originally allowed by IRS after return processing.

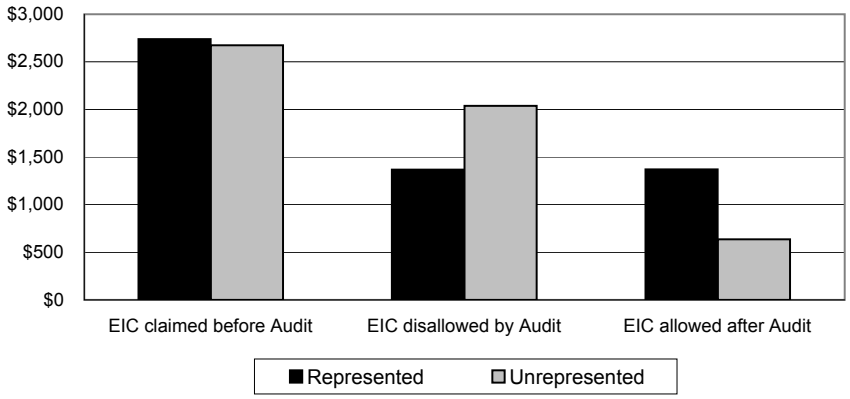
Table 6. EIC Amount Before and After Audit[33]

Average EIC Amount:	Represented	Unrepresented	Difference (Rep.—Unrep.)
Before audit	$2,740	$2,674	$66
Disallowed during audit	$1,369	$2,038	-$669
After audit	$1,371	$636	$735

Source: IRTF TY2002 and CAF for TY2002

In summary, represented taxpayers retain more of their EIC in both percentage and absolute dollars over taxpayers without representation.

Figure 3. Amount of EIC Before and After Audit



Source: IRTF TY2002 and CAF for TY2002

Objective 3: Determine if the tax recommended for taxpayers with representation in EIC audits is less than the tax recommended for taxpayers without representation in EIC audits.[34]

Fewer represented taxpayers owe additional tax.

The prior findings focus on the impact of representation on the EIC. There may be other issues addressed during an audit that can offset adjustments to the EIC. In this section, we investigate the impact of representation on the net tax resulting from the audit.

[33] The difference between represented and unrepresented taxpayers is statistically significant at level .0001 (one-sided t-test).

[34] This objective focuses only on tax change which is separate from changes in refundable credits such as EIC.

The average amount of additional tax due after audit for both represented and unrepresented taxpayers is similar, as shown in Table 7 below. Nevertheless, there are significant differences in the percentage of taxpayers within these two groups who actually owe additional tax. Over 60 percent of the represented group owe no additional tax, while over 54 percent of unrepresented taxpayers owe additional tax at the conclusion of the audit.

Table 7. Tax Change During Audit[35]

Percentage of taxpayers:	Represented	Unrepresented
whose tax increased during audit	38.2%	54.4%
with no change in tax during audit	56.0%	36.5%
whose tax decreased (refund) during audit	5.8%	9.1%
Average Tax Change (increase) during audit	$291	$304

Source: IRTF TY2002 and CAF for TY2002

The one positive result for unrepresented taxpayers is that 9.1 percent of them received a reduction in tax due versus 5.8 percent for represented taxpayers. However, this needs to be considered in combination with the greater share of unrepresented taxpayers (54.4 percent) who pay additional tax.

Objective 4: Determine the extent of the effect of representation on the outcome of EIC audits.

Even when attempting to control for self-selection bias, representation was still found to be a significant factor in the audit outcome.

The prior findings show significant differences between represented and unrepresented taxpayers. Nevertheless, the possibility exists that these differences are the result of inherent differences between those taxpayers who seek representation versus those taxpayers who do not.

Our goal was to determine whether or not representation is a factor in EIC loss due to audit. Initially, we planned to develop a linear regression model to estimate EIC loss using representation as a covariate as well as several other factors and thus test our hypothesis. Our data provided us with two hurdles to overcome before we could test our hypothesis. One problem was that the data are not derived from a randomized trial. Taxpayers were not provided the treatment of representation randomly; rather, taxpayers chose to be represented on their own. This could introduce self-selection bias and result in the represented group having different characteristics than

[35] The difference between represented and unrepresented taxpayers is statistically significant at level .0001 (one-sided t-test).

the unrepresented, thus introducing bias into estimators of our regression. Second, our data were not normally distributed with respect to EIC loss. In particular, we had a large number of taxpayers who incurred no loss, and we felt that we may not meet the assumptions of linear regression. We did not want to delete taxpayers with no loss in EIC after audit since no loss is a result of an audit. Considering both issues, we determined that we needed to account for selection bias, and it would be unlikely that we could successfully use linear techniques. We chose to use propensity score matching to account for selection bias and use a logistic regression model to estimate the probability of EIC loss or no EIC loss using represented as a covariate and using the matched data.[36]

After examining whether representation was a factor in EIC loss, we also looked at whether or not representation is a factor in the amount of EIC loss for taxpayers who lose EIC due to audit using a linear regression. Removing taxpayers with no EIC loss helped our data conform to the assumptions of linear regression. The linear regression gives us a sense of the magnitude of representation's effect on EIC loss for those who lose EIC.

Propensity score matching will help even out the differences in characteristics between the represented and unrepresented. A brief summary of our methods follows in the body of this report to better explain the results; however, a more complete technical explanation is contained in Appendix 1.

The first step is to estimate the propensity or probability of a taxpayer to be represented using the data we have as covariates in a logistic regression. The next step is to segment the data into deciles of propensity score. Each decile contains taxpayers who are similar with respect to the covariates. If our logistic regression is viable, the lower deciles will have a smaller number of represented taxpayers, and the higher deciles will have a larger number of represented taxpayers. In order to even out the number of represented and unrepresented in each decile, we randomly sampled the majority group within each decile to match the number of the minority group. The data remaining after sampling are our matched data that we use to develop regressions to determine if representation is a factor in EIC loss due to audit. Statistical tests are used to examine the extent to which the above matching procedures resulted in samples of represented and unrepresented taxpayers who were more comparable in terms of baseline characteristics.

As stated above, the first step of the propensity score matching is to develop propensity scores for representation. The percentage of those taxpayers with representation in a dataset is rather small at around 4 percent. Any

[36] For specific details regarding the utilization of propensity score matching with these data and for greater detail on the subsequent logistic regression models, see Appendix 1.

model developed with this data set would grossly underestimate the probability of representation. Therefore, before we developed the logistic regression for estimating the probability of representation, we randomly sampled the unrepresented so that the number of unrepresented roughly equaled the number of represented (50.3 percent versus 49.7 percent, respectively). We eliminated audit cases closed as "no shows." Since audits for unrepresented taxpayers were nearly twice as likely to be closed without taxpayer response and since, by definition, the presence or absence of a representative is irrelevant in these cases, we felt removing these cases resulted in a fairer comparison of the two groups. We also used only cases with a Dependent Database (DDb) score for our sample to better account for potential compliance differences between the groups.[37] Our logistic regression model for estimating the propensity of representation used paid preparer, Schedule C, age, form type, gender, adjusted gross income (AGI) bands, and EIC criteria as the covariates. A 50-percent sample was used to train the model.[38] The overall accuracy of this model was around 64 percent. The results of the propensity score matching can be seen in Table A-1, Appendix 1.

After completing the first logistic regression model to estimate the propensity of representation, we developed a second logistic regression model to estimate the effect of representation on EIC loss. This model utilized representation, DDb score Married Filing Joint (MFJ) filing status, and specific EIC project codes as covariates. A 50-percent sample was used to train the model. The overall accuracy of the model is 66 percent. We determined that representation is a factor in predicting the loss of EIC, and we estimate that, if a taxpayer is represented, the odds increase by a factor of over two that he or she will not lose EIC during audit. As a comparison, for DDb score divided by 10, we estimate that, for every 10-point increase in DDb score, the odds of some EIC loss increase by a factor of 1.12. See Table A-2, Appendix 1.

As stated above, after we determined whether representation is a factor in EIC loss, we looked at the relationship of representation to the dollar amount EIC lost for taxpayers who lose some EIC due to audit. Eliminating the taxpayers with no EIC loss normalized the distribution of EIC loss enough to warrant an attempt at a linear regression. Represented, DDb score, category of EIC audit issue, gender, AGI, and number of qualifying children were used to estimate EIC lost during audit. Represented is signifi-

[37] The Dependent Database (DDb) is a tool that identifies noncompliant Earned Income Tax Credit (EITC) and dependent issues through the use of internal and external data elements and provides the ability to freeze refunds. The database is rule-driven. If a rule condition is met as returns are processed through the DDb rule filtering process, the rule "fires," and the return is flagged for examination. Current procedures score the majority of all EIC returns; however, for Tax Year 2002, fewer returns were scored (about 60 percent of our study cases).

[38] The overall accuracy of the regression is 64 percent with c statistic of .644 (area under the ROC).

cant and has a coefficient estimate of $127. Noteworthy, DDb score, AGI, and number of qualifying children were not significant in the model. See Appendix 1 for details of the linear regression.

Objective 5: Compare return and other demographic characteristics of the EIC audit population with representation to those without representation.

Although some notable demographic differences exist between taxpayers who use and do not use representation during EIC audits, the positive effect of representation is still evident.

This objective was originally envisioned before the more rigorous approach to addressing self-selection bias was adopted for the preceding objective. The plan was to compare the demographic characteristics of two groups to determine the potential self-selection bias. The most significant differences between the taxpayers represented and unrepresented during audit are shown in Table 10.

Table 10. Most Significantly Different Characteristics Between Represented and Unrepresented (During Audit) EIC Taxpayers

Characteristic	Represented	Unrepresented
Married Filing Joint FS	18.8%	8.7%
Head of Household FS	70.8%	78.9%
Male filers	57.0%	66.3%
Used paid preparer on return	86.0%	74.4%
Form 1040	69.6%	56.7%
Form 1040A	30.4%	43.2%
Schedule C	50.0%	36.3%
Two qualifying EIC children	60.1%	49.6%
No Show/No Response	19.7%	37.7%

Source: IRTF TY2002 and CAF for TY2002, and SSA Data Master-1file 2006

Three of these factors (gender, paid preparer, and form type) were shown to be significant in the propensity score matching. Additionally, a

composite of two more of the above variables (number of children and the married filing joint filing status) was also shown to be a significant factor in the propensity scoring. Two of these factors (presence of Schedule C and the MFJ filing status) were shown to be significant in the analysis of the selection bias in the preceding objective. Not unexpectedly, Table 10 also indicates that represented taxpayers are significantly more likely to respond to the audit notice than unrepresented taxpayers.

In general terms, taxpayers with representation are more likely to use the MFJ filing status, use a paid preparer, file Form 1040 with a Schedule C, have a balance due at the time of filing, and have two qualifying children for EIC. Taxpayers without representation during an audit are more likely to be filing as Head of Household (HOH), filing Form 1040A, claiming the maximum amount of EIC, and not responding to IRS notices and/or not showing up for the audit.

To further explore the relationship between the factors in Table 10 and the effect of representation, we created several tables which cross-tabulate these factors with the audit results.

Table 11. Preparer Type: Average EIC Amount Disallowed (Reduced) After Audit[39]

	Represented	Unrepresented
Overall Average[40]	$1,369	$2,038
Self-Prepared Returns	$1,387	$2,056
Paid Preparer Returns	$1,366	$2,032

Source: IRTF TY2002 and CAF for TY2002

Table 11 shows that paid preparer returns retained a slightly higher amount of the EIC after audit than self-prepared returns; however, the differential between represented and unrepresented taxpayers remains. Some might argue that represented taxpayers retain more of their EIC because their return was more accurate in the first place because the return was completed by a paid preparer.[41] However, the data show that the impact of having a representative is larger than whether a paid preparer was used on the original return.

[39] The difference between represented and unrepresented taxpayers is statistically significant at level .0001 (one-sided t-test).

[40] Overall, represented taxpayers had 50 percent of their EIC disallowed, compared to 76 percent for unrepresented taxpayers.

[41] The EIC under review during an audit is the net amount after math error processing that occurs when the tax return is filed.

Table 12. Number of Qualifying Children—Average Amount of EIC Disallowed

Qualifying Children	Average EIC Disallowed	
	Represented	Unrepresented
No Children	$26	$92
One Child	$1,105	$1,739
Two Children	$1,546	$2,353

Source: IRTF TY2002 and CAF for TY2002

Table 12 shows that represented taxpayers fare better across each possible number of claimed eligible EIC children. As expected, more EIC is disallowed for returns claiming two children; however, represented taxpayers claiming two EIC children actually lose less EIC, on average, than unrepresented taxpayers claiming only one child.

Table 13. Gender—Average Amount of EIC Disallowed

Gender	Average EIC Disallowed	
	Represented	Unrepresented
Female	$1,279	$1,858
Married, Filing Joint Return	$900	$1,100
Male	$1,562	$2,229

Source: IRTF TY2002 and CAF for TY2002 and SSA Data Master-1 file 2006.

Males typically lose the largest amount of EIC during audit, while joint filers lose the least. Represented males; however, lose less EIC on the average than unrepresented females.

Table 14. Filing Status—Average Amount of EIC Disallowed

Filing Status	Average EIC Disallowed	
	Represented	Unrepresented
Single	$1,520	$2,120
Married, Filing Joint Return	$900	$1,100
Head of Household	$1,472	$2,129
Qualifying Widower	$1,111	$1,396

Source: IRTF TY2002 and CAF for TY2002

Single and head of household filers have the largest EIC change, while married filing joint taxpayers have the lowest average audit change. Represented taxpayers fared better than their unrepresented counterparts in each filing status.

Table 15. Adjusted Gross Income—Average Amount of EIC Disallowed

Adjusted Gross Income	Average EIC Disallowed	
	Represented	Unrepresented
Less than $5,000	$810	$1,051
$5,000 to $9,999	$1,496	$2,189
$10,000 to $14,999	$1,716	$2,495
$15,000 to $19,999	$1,378	$1,990
$20,000 to $24,999	$935	$1,261
$25,000 to $29,999	$433	$548
$30,000 and Over	$167	$209

Source: IRTF TY2002 and CAF for TY2002

Those AGI categories which correspond to the highest EIC entitlement are the same categories that generate the highest average EIC Change. Again, represented taxpayers have a lower EIC audit change amount across each AGI range.

Table 16. Schedule C—Average Amount of EIC Disallowed

Schedule C Filed	Average EIC Disallowed	
	Represented	Unrepresented
No Schedule C	$1,306	$1,986
Schedule C	$1,432	$2,129

Source: IRTF TY2002 and CAF for TY2002

Taxpayers represented in an EIC audit are much more likely to have filed a Schedule C. These represented taxpayers also have a significantly smaller average EIC change.

Conclusions and Recommendations

Conclusions

- Taxpayers with representation are twice as likely to be found eligible for the EIC as taxpayers without representation during the audit process.
- Over one-half of all taxpayers with representation emerged from audits with their full EIC intact, whereas less than 1 in 4 taxpayers without representation kept their full EIC.
- Taxpayers without representation were more likely to end up owing additional tax than taxpayers with representation (54 percent versus 38 percent).
- Taxpayers without representation were less likely to end up with no change in tax after audit than taxpayers with representation (37 percent versus 56 percent).
- Even when controlling for self-selection bias, taxpayers without representation are still two times more likely to have their EIC reduced than taxpayers with representation.
- Although some significant differences exist between taxpayers with and without representation during EIC audits, these differences do not overcome the positive effect of representation on the audit outcome.

Recommendations

TAS's initial recommendation is that this study be replicated on a more recent tax year to see if the presence of a taxpayer representative during an EIC audit continues to have a significant positive impact on the outcome.

Appendix 1

Propensity score matching is a two stage process.[42] In the first stage, the propensity score, which is the likelihood (or propensity) of a case being in the represented group, is estimated for each case through a logistic regression model. The represented and unrepresented groups are each then sampled to identify subsamples with similar distributions of this estimated score.

[42] See Rosebaum, P. R. and D. B. Rubin, Biometrika 1983:70, pp.41-55.

As shown by Rosenbaum and Rubin, such matching results in subsamples of the study and control group with similar distributions of observed risk factors. Cochran states that creating five strata removes 90 percent of the bias due to the stratifying variable or covariate, and Rosenbaum and Rubin claim that stratification on propensity score removes even more than this in each covariate used in the propensity model.[43] While researchers typically only check to see that the matched subsamples have similar means for all-important risk factors, more sophisticated checks on the comparability of their multivariate distributions can also be done. In the second stage, standard analytic techniques are used to fit a response model to the matched subsamples and, ultimately, to estimate the effect of the representation on the outcome. The propensity score method typically has less precision, due to reduced sample sizes, but this limitation is generally of less concern than the worry over possible bias from a misspecified model. It should also be noted that propensity score matching does not address or resolve problems due to imbalances in unmeasured factors.[44]

One may consider whether the model used to estimate the propensity score might itself be misspecified, introducing a new set of problems for the analysis. Drake used simulations to compare consequences of misspecifications of the propensity score to those of misspecified response models.[45] She concluded that the propensity score ‘‘seems preferable when considering model misspecifications in the response model, particularly so because an incorrect propensity score model has smaller bias’’ and ‘‘generally, the simulations seem to indicate that the value of the propensity score lies primarily in guarding against model misspecifications.’’[46]

In the first stage of our model, we developed a logistic regression using paid preparer, Schedule C, age, Form 1040, gender, AGI bands, and EIC criteria to predict the propensity to be represented. A 50-percent sample was used to train the model. The overall accuracy of the regression is 64 percent with c statistic of .644 (area under the ROC). We split the data into deciles based on the propensity scores. To develop the matched samples, within each decile, a random sample of the larger group (represented or unrepre-

[43] See Cochran, W. G., “The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies,” Biometrics, 24, pp. 205-213.

[44] Whole paragraph footnoted to M. A. Posner, A. S. Ash, K. M. Freund, M. A. Moskowitz, and M. Swartz, “Comparing Standard Regression, Propensity Score Matching, and Instrument Variable Matching Methods for Determining the Influence of Mammograms and Stage Diagnosis.

[45] See Drake, C. (1993), “Effects of Misspecification of the Propensity Score on Estimation of Treatment Effect”, Biometrics, 49, pp. 1231-1236.

[46] Whole paragraph footnoted to M. A. Posner, A. S. Ash, K. M. Freund, M. A. Moskowitz, and M. Swartz, “Comparing Standard Regression, Propensity Score Matching, and Instrument Variable Matching Methods for Determining the Influence of Mammograms and Stage Diagnosis.”

sented) was taken to get the same number in the smaller group. The matched subsamples were then combined to create the matched dataset which we used for the second step. As depicted in the following table, to examine the extent to which the above matching procedures resulted in samples of users and nonusers more comparable in terms of baseline characteristics, statistical tests were used; p-values for chi-squared tests of independence were calculated for categorical risk factors, and p-values for independent sample t-tests for equivalence of population means were calculated for continuous risk factors.

Table A-1

	Prematching			Postmatching		
Predicted represented	**Rep**	**Unrep**		**Rep**	**Unrep**	
Decile 1	229	678		229	229	
Decile 2	273	617		273	273	
Decile 3	345	563		345	345	
Decile 4	356	544		356	356	
Decile 5	422	484		422	422	
Decile 6	460	443		443	443	
Decile 7	523	354		354	354	
Decile 8	577	337		337	337	
Decile 9	623	296		296	296	
Decile 10	677	210		210	210	
Total	4,485	4,526	9,011	3,265	3,265	6,530
Average Primary Age	37.2	34.4	p<.001	35.56	35.73	p=.495
Age Cat						
17 and Under	0.2%	0.5%		0.3%	0.3%	
18-29	23.8%	36.5%		29.3%	31.6%	
30-39	35.9%	32.2%		36.9%	33.0%	
40-49	29.6%	22.6%		25.9%	25.0%	
50-64	9.5%	7.6%		7.0%	9.3%	
65 and Over	1.0%	0.6%	p<.001	0.7%	0.7%	p=.50
AGI Bands						
<= 9740.00	24.5%	30.8%		29.0%	28.4%	
9740.01 - 13172.00	31.3%	32.3%		31.8%	32.2%	
13172.01 - 18136.00	27.5%	23.8%		25.0%	25.7%	
18136.01+	16.6%	13.1%	p<.001	14.2%	13.7%	p=.82
Paid Preparer						
No	33.8%	16.4%		23.9%	22.1%	
Yes	66.2%	83.6%	p<.001	76.1%	77.9%	p=.09

	Prematching			Postmatching		
Predicted represented	**Rep**	**Unrep**		**Rep**	**Unrep**	
Schedule C						
No Schedule C	59.4%	40.6%		49.9%	49.3%	
Schedule C	40.6%	59.4%	p<.001	50.1%	50.7%	p=.62
Criteria 2 (MFJ, 1Q child, AGI<3,178)						
Not Criteria 2	96.2%	90.0%		95.0%	94.5%	
Criteria 2	3.8%	10.0%	p<.001	5.0%	5.5%	p=.374
Gender						
F	22.3%	18.9%		22.0%	21.6%	
J	4.6%	11.6%		5.9%	7.1%	
M	73.1%	69.5%	p<.001	72.1%	71.3%	p=.162
Avg. EIC Disallowed	$2,031	$1,317.00	p<.001	$1,994	$1,357	p< .001
Avg. DDb Score	83.6	70.3	p<.001	81	72	p<.001

Source: IRTF TY2002 and CAF for TY2002 & DDb

For the second stage of our model using the matched data, we developed a logistic regression model using representation, Dependent Database (DDb) score, Married Filing Joint (MFJ) filing status, and specific EIC project codes to predict the propensity of EIC lost or no change or positive change in EIC.[47, 48] A 50-percent sample was used to train the model. The overall accuracy of the model is 66 percent with c statistic .609. Table A-2 below shows the coefficients and odds ratio (EXP(B)).

Table A-2: Significant Factors Determining Likelihood of Retaining EIC After Audit

Factor	Coefficient	Odds Ratio
Representation	-.745	0.47
DDb Score (divided by 10)	0.12	1.12
Audit Project Code 623	.510	1.67
Married Filing Joint Filing Status	-.904	0.41
AGI Between $9,470 and $13,172	-.118	0.82
Audit Project Code 624	-.333	0.72
Constant	.010	1.01

Source: IRTF TY2002 and CAF for TY2002 &DDb

[47] We limited our regression analysis to cases with a DDb score to better account for compliance differences between represented and unrepresented taxpayers.

[48] A score assigned to returns depending on specific taxpayer and return circumstances.

The estimated coefficient (B) of representation is negative with an odds ratio of 0.47. So, it is estimated that, if a taxpayer is represented, the odds increase by a factor of two (1/.47) that he or she will not lose EIC during audit. As a comparison, DDb score divided by 10 has an estimated positive coefficient with an odds ratio of 1.12, so that it is estimated that, for every 10-point increase in DDb score, a taxpayer's odds of EIC lost is increased by factor of 1.12, which is less than if represented.

The model predicts that a taxpayer without representation has more than twice the odds of a represented taxpayer to have the EIC reduced. This is true controlling for filing status, DDb score, adjusted gross income, and type of audit issue.

After examining the effect of representation on the likelihood of retaining all EIC after audit, we then explored a linear regression model to estimate the effect of representation on the amount of EIC lost after audit. We developed this model from the same matched sample utilized for our prior logistic regression model. Eliminating the no change EIC cases eliminated the distribution spike at zero in EIC loss. Represented, DDb score, category of EIC audit issue, gender, AGI, and number of qualifying children were used to estimate EIC lost during audit. DDb score, AGI, and number of qualifying children were not significant in the model.

Table A-3 below shows the unstandardized coefficients representing the effect of significant factors in the model.[49]

Table A-3: Significant Linear Regression Factors Estimating Amount of EIC Lost During Audit

Factor	Coefficient
Earned Income Credit (IRS Computation)	.660
Representation	-126.89
AGI Between $9,470 and $13,172	368.10
AGI Less Than $9,470	301.51
MFJ Fil. Stat; 2 Qual. Ch., AGI> 34K	153.14
Female Gender	-96.41
Constant	541.2

Source: IRTF TY2002 and CAF for TY2002 & DDb

[49] Significance of factors was determined at the .05 level. The adjusted R-Square was .52.