

## EITC Interactive: User Guide and Data Dictionary

This guide answers general questions about the operation of the [interactive website](#) and the information it provides, as well as specific questions users might ask about how to use some of these data.

**Please note that the data have changed substantially between Tax Year 2013 and Tax Year 2014, due to changes in IRS reporting standards. See the [section](#) on changes to the data for full details.**

For more information about the database and using the data, contact [Cecile Murray](#).

### General Questions

- [What's new in this installment of the website?](#)
- [What do the different geography types mean?](#)
- [How are these data derived?](#)
- [What do the column headers in the data represent?](#)
- [Are additional data available?](#)

### EITC-Specific Questions

- [How many people in my community benefit from the EITC?](#)
- [What's the typical value of the EITC that people in my community receive?](#)
- [What do the RAL and RAC variables mean? How can I use them?](#)
- [How many people in my community are eligible for the EITC, but don't receive it? What's the amount that is "left on the table" as a result?](#)
- [Can I use these data to characterize the total economic impact that EITC dollars have in my local economy?](#)

---

## GENERAL QUESTIONS

### WHAT'S NEW IN THIS INSTALLMENT OF THE WEBSITE?

This version of the EITC Interactive contains data for Tax Year 2014, the most recent year of data available. (Data from Tax Years 1997 through 2013 are available upon request by e-mailing [Cecile Murray](#).)

All EITC Interactive data are derived from the Internal Revenue Service's Stakeholder Partnerships, Education, and Communication (IRS-SPEC) Return Information Database, compiled by the IRS Wage and Investment Research Unit. The IRS-SPEC database presents data at the ZIP code level.

**Note: The IRS imposed much more stringent data sanitization requirements on the IRS-SPEC ZIP code data in Tax Year 2014 compared to previous years, which has led to significant changes in what is available on EITC Interactive.**

Specifically, the IRS-SPEC database for Tax Year 2014:

- Only includes information on returns with adjusted gross incomes (AGI) under \$60,000;
- Only includes data for ZIP codes that have more than 100 filers with AGIs under \$60,000. All other ZIP codes are suppressed;
- Only presents breakouts of return characteristics (e.g., credits/deductions claimed, method of return preparation, AGI category) if there are more than 20 filers in a given category. (The previous suppression threshold was 10 returns.)

As a result of these significant changes to the IRS-SPEC data reporting requirements, this version of the EITC Interactive website:

- **No longer includes information on the number and characteristics of all tax returns.** Only information on returns claiming the EITC is currently available through the application.
- **No longer includes data on EITC filers by qualifying children or the ITIN population.** Data on EITC receipt by the number of qualifying children was not included in the Tax Year 2014 IRS-SPEC database given the extensive level of suppression. Data on the number of ITIN filers is no longer available because EITC Interactive now only includes data on EITC filers, who must have a valid social security number to claim the credit.
- **The higher incidence of suppression means any comparisons over time and across places should be made with caution and may not be appropriate for all variables or geography types.** Some variables are more affected by suppression than others (e.g., filers who did not file a return in the previous year, returns prepared by volunteers, returns by AGI category). Rural areas and lower levels of geography are also more likely to be affected by suppressions.

[\[Return to top\]](#)

## WHAT DO THE DIFFERENT GEOGRAPHY TYPES MEAN?

EITC Interactive allows users to request data for a variety of different communities. ZIP code-level IRS data have been split up and aggregated to create larger geographic units, including cities, counties, metro areas, states, state legislative districts, and congressional districts. Each geography type is briefly described below.

**ZIP Codes** are designated by the U.S. Postal Service and are the smallest areas for which the IRS reports tax return data. ZIP codes provide the building blocks for all other geography types on the EITC Interactive site. When downloading the state-by state ZIP Code files, please note that the data is displayed only for the portion of the ZIP Code that is within that state (ZIP Codes can, in rare instances, cross state boundaries).

**Places (Cities/Towns)** include incorporated places, such as cities, towns, and villages, as well as census-designated places, which are unincorporated areas delineated by the U.S. Census Bureau for statistical purposes.

**Counties** represent the primary legal subdivision of most states. Exceptions include Alaska and Louisiana, which are divided into boroughs and parishes, respectively.

**Metropolitan Areas** are composed of counties. Metro names in this version of the site are based on the [2013 metropolitan statistical area delineations](#) issued by the U.S. Office of Management and Budget. Because metropolitan areas can cross state boundaries, data for metropolitan areas are only available in one national file.

**State** files provide data on all ZIP codes in the state including their associated cities, counties, and metro areas.

**States Legislative Districts** reflect boundaries in effect as of the 2012 elections. EITC Interactive includes data for lower chamber (generally House) and upper chamber (generally Senate) districts in all 50 states and the District of Columbia. Note that Nebraska has a unicameral system. Data for Nebraska can be accessed through either the "lower chamber" or "upper chamber" geography type.

**Congressional Districts** represent the 114<sup>th</sup> United States Congress. Congressional district boundaries did not change between the 113<sup>th</sup> and 114<sup>th</sup> Congress.

In the current version of the EITC Interactive application, users directly download state-level files that contain all available variables summarized to the requested level of geography. In addition, users can download national level files for each geography type. (Users who are interested in seeing the underlying ZIP code allocation can e-mail [Cecile Murray](#).)

[\[Return to top\]](#)

## HOW ARE THESE DATA DERIVED?

The ZIP code-level data that the IRS provides to Brookings contain county, state, and city identifiers. However, the IRS assigns ZIP codes to cities and towns based primarily on information from the U.S. Postal Service, which associates a ZIP code with the name of the city or town nearest to its post office location. In many instances, this does not reflect the location of the bulk of the ZIP code itself, because ZIP codes do not conform to municipal boundaries. In addition, ZIP codes do not always conform to county boundaries. In cases where ZIP codes cross multiple counties, the IRS uses information from the U.S. Postal Service to identify the primary county.

Rather than using the IRS identifiers, to assign ZIP codes to cities and counties, we used Geographic Information Systems (GIS) and statistical software to identify where ZIP codes are located. For ZIP codes that cross city and/or county boundaries, we used Census block-level data (the smallest units for which the Census Bureau tabulates data), along with census places and ZIP code boundaries, to calculate the proportion of the ZIP code's households that lie within each geography. We undertake the same process to assign ZIP codes and partial ZIP codes to state legislative and congressional districts. Return and dollar amount are estimated by allocating ZIP code totals based on the percentage of the ZIP code's households that fall within the geography's borders.

For the sake of confidentiality, the IRS suppresses return counts of less than 20. We impute suppressed totals at the ZIP code level for the following variables: eic, ctc, actc, ref, and bal; however, all other variables may be subject to data suppression.

**Note: Because of the estimation techniques employed in assigning ZIP codes to cities counties, state legislative districts, and congressional districts, the data displayed here will differ from geography totals obtained directly from the IRS files. In addition, some data displayed will include return counts of less than 20. These represent estimates only. Small counts should be interpreted only in conjunction with data from other geographies (e.g., at the place, county, or district levels) and not as stand-alone entities.**

[\[Return to top\]](#)

## WHAT DO THE COLUMN HEADERS IN THE DATA REPRESENT?

The website returns data to the user with a series of abbreviated column header names.

Each column header contains the prefix "e," which refers to **"EITC Returns Only."**

The root of the variable name refers to the descriptions below.

VARIABLE NAME (ROOT)	VARIABLE DESCRIPTION
returns	Total number of returns
new	Total number of returns where the taxpayer did not file in the previous year
eic	Total number of returns receiving the Earned Income Tax Credit (EITC)
eicam	Sum of EITC received
ctc	Total number of returns receiving the Child Tax Credit
ctcam	Sum of Child Tax Credit received
actc	Total number of returns receiving the refundable portion of the Child Tax Credit
actcam	Sum of the refundable Child Tax Credit received
cdctc	Total number of returns filing Form 2441 (Child and Dependent Care Expenses)
edcr	Total number of returns filing Form 8863 (Education Credits)
sld	Total number of returns receiving a deduction for payment of Student Loan Interest <b>Note:</b> This is distinct from the tuition and fees deduction and any education credits.
ref	Total number of returns receiving a refund

---

VARIABLE NAME (ROOT)	VARIABLE DESCRIPTION
refam	Sum of refunds received
bal	Total number of returns with a balance due after remittance
balam	Sum of balance due after remittance
dirdp	Total number of returns receiving direct deposit of refund <b>Note:</b> Refund anticipation products are counted in this variable because they direct refunds to temporary bank accounts through direct deposit.
ral	Total number of returns requesting a Refund Anticipation Loan (RAL) <b>Note:</b> Beginning in Tax Year 2010, IRS removed the "debt indicator," an indication of whether the taxpayer has outstanding debt. As a result, the number returns requesting a RAL dramatically decreased.
rac	Total number of returns requesting a Refund Anticipation Check (RAC)
self	Total number of returns that were prepared by taxpayer <b>Note:</b> This category includes filers who purchased software to prepare and file returns from home, Free File Alliance filers, and some volunteer-facilitated self-preparation
paid	Total number of returns prepared by a paid preparer
vol	Total number of returns prepared by volunteer organizations (VITA, Military VITA and TCE)
1040_	Total number of returns filed on Form 1040

---

VARIABLE NAME (ROOT)	VARIABLE DESCRIPTION
1040a	Total number of returns filed on Form 1040A
1040z	Total number of returns filed on Form 1040EZ
cef	Total number of returns that filed one or more of the following schedules: Schedule C (Profit or Loss from a Business); Schedule E (Supplemental Income and Loss); Schedule F (Profit or Loss from Farming)
agi0_	Total number of returns with Adjusted Gross Income less than \$5,000
agi5_	Total number of returns with Adjusted Gross Income from \$5,000 to \$9,999
agi10_	Total number of returns with Adjusted Gross Income from \$10,000 to \$14,999
agi15_	Total number of returns with Adjusted Gross Income from \$15,000 to \$19,999
agi20_	Total number of returns with Adjusted Gross Income from \$20,000 to \$24,999
agi25_	Total number of returns with Adjusted Gross Income from \$25,000 to \$29,999
agi30_	Total number of returns with Adjusted Gross Income from \$30,000 to \$34,999
agi35_	Total number of returns with Adjusted Gross Income from \$35,000 to \$39,999
agi40_	Total number of returns with Adjusted Gross Income from \$40,000 to \$49,999

---

VARIABLE NAME (ROOT)	VARIABLE DESCRIPTION
-------------------------	----------------------

agi50_	Total number of returns with Adjusted Gross Income from \$50,000 to \$59,999
agi60_	Total number of returns with Adjusted Gross Income from \$60,000 to \$74,999
agi75_	Total number of returns with Adjusted Gross Income from \$75,000 to \$99,999
agi1k_	Total number of returns with Adjusted Gross Income greater than or equal to \$100,000

[\[Return to top\]](#)

#### ARE ADDITIONAL DATA AVAILABLE?

The data available on this site reflect some, but not all, of the data published by IRS-SPEC in its Tax Return Information Database. Users interested in viewing the full complement of data available should contact their [local SPEC Territory Managers](#) to obtain a copy of the database.

[\[Return to top\]](#)

---

#### EITC-SPECIFIC QUESTIONS

##### HOW MANY PEOPLE IN MY COMMUNITY BENEFIT FROM THE EITC?

The total number of tax filers claiming the EITC is given by the **eic** field. In general, filers receive this and other tax credits in the year following the tax year for which they file. For instance, most filers who claimed the EITC for Tax Year 2014 received the credit upon filing returns in early 2015.

Because the total tax filing universe is not available in the IRS-SPEC or EITC Interactive Tax Year 2014 data, the data cannot be used to compute an important related measure: the proportion of tax filers in a community who receive the EITC.

However, [IRS-Statistics of Income \(SOI\)](#) does tabulate data on total tax filers and EITC filers at the ZIP code, county, and state level, which would allow users to calculate the share of filers claiming the EITC for those geography types.

Note that SOI and SPEC estimates differ somewhat due to slight differences in the methods used to compile the data, so they should not be directly combined to create proportions or other tabulations. But users may wish to make use of both data sources to provide a fuller picture of tax filers in their community than they otherwise could. For instance, the SOI data offers a look at the total tax filing universe and some key breakouts (e.g., returns claiming the Child Tax Credit, returns prepared with volunteer assistance), but does not provide the detailed information on characteristics of the EITC

population available through the SPEC data (e.g., the number and share of EITC filers claiming the Child Tax Credit or using volunteer tax preparation services).

[\[Return to top\]](#)

### **WHAT'S THE TYPICAL VALUE OF THE EITC THAT PEOPLE IN MY COMMUNITY RECEIVE?**

Depending on their income and family structure, tax filers may claim an EITC the value of which could range from \$1 to over \$6,000. For your community, in any given year you can find the average credit that EITC filers received by dividing the **eicam** field by the **eic** field. In Tax Year 2014, the average credit received among EITC filers nationwide was \$2,449.

Two primary factors influence the average EITC claim in your community:

- Childless workers with very low incomes (under \$14,590 for single filers in Tax Year 2014) are eligible for a much smaller credit -- up to \$496 in 2014 -- than workers with children, whose maximum credits in 2014 are \$3,305 (for families with one qualifying child), \$5,460 (for families with two or more qualifying children), and \$6,143 (for families with three or more qualifying children). If a higher-than-average share of EITC recipients in your community are childless workers, the average credit amount will likely be smaller. Nationally, about 80 percent of EITC recipients claim the credit for families with qualifying children.
- The wages earned by low-income families influence the credit amount for which they are eligible. Nationally, over half of filers who received the EITC in Tax Year 2014 had adjusted gross incomes above \$15,000. For every additional dollar families with children earned above \$17,830 (\$23,260 for married families), the amount of credit they received decreased. Thus, EITC filers in higher-wage, higher cost-of-living areas tend to receive smaller credits; conversely, average credits in lower-wage areas tend to be larger.

Note that EITC dollars claimed and tax refunds received are not equivalent. Some EITC dollars (14 percent nationwide) offset other taxes that families owe, and thus do not translate directly into refund dollars. Additionally, families may claim other credits (like the Child Tax Credit, the Child and Dependent Care Tax Credit, and Education Credits) that add to their refunds, and some refund dollars represent taxes that were over-withheld over the course of the year. For most low-income families who receive tax refunds, however, the EITC makes up the largest part of those refunds— and is thus the most important part of the federal tax code for many low-income communities. Users can calculate the relative contribution of the EITC to low-income taxpayers' refunds by comparing the **eeicam** and **erefam** variables in the selected geography.

[\[Return to top\]](#)

### **WHAT ARE THE RAL AND RAC VARIABLES? HOW CAN I USE THEM?**

RAL is an abbreviation for refund anticipation loan while RAC stands for refund anticipation check, products sold by most commercial tax preparers. Low-income taxpayers who claim the EITC represent the majority of the marketplace for both products.

By purchasing a RAL, the tax filer assigns the proceeds of his/her tax refund to the preparer's bank partner, and the preparer arranges a loan for the taxpayer in the amount of his/her refund, net of fees for tax preparation and the loan itself. The bank makes the loan available to the taxpayer within 1-2 days, and the IRS typically delivers the taxpayer's refund to the bank within about 10 days. For this short-term loan, the taxpayer often pays fees in excess of \$100 (in addition to the fees they pay to have their taxes prepared and filed), and incurs an implicit annual interest rate on the loan of 250 percent or higher.

**Starting in Tax Year 2010, the IRS stopped providing commercial tax preparers with a "debt indicator," an indication of whether the taxpayer has outstanding debt. As a result, [research has shown](#) the number of returns**



**requesting a RAL has dramatically decreased.** In Tax Year 2007, 38 percent of EITC recipients using paid preparers requested RALs. By 2010, this percentage dropped to just over 5 percent.

As the use of RALs has declined, an increasing number of tax filers have requested RACs. With a RAC, instead of issuing a loan within 1-2 days, the bank opens a temporary bank account into which the IRS direct deposits the refund check. After the refund is deposited, the bank issues the consumer a paper check or prepaid debit card with the RAC proceeds and closes the temporary account. This process usually takes 7-15 days. Between Tax Year 2007 and 2010, the percentage of EITC recipients using paid preparers who requested RACs increased from 26 percent to 56 percent.

The **ral** variable in the database represents the number of returns for which the taxpayer requested a RAL. To determine the proportion of filers in your community who requested RALs, divide the **ral** field by the **ref** field for the tax year selected. To determine the proportion of filers requesting RACs in a given year, divide the **rac** field by the **ref** variable.

To estimate how much money EITC filers in your community are spending on RALs and RACs, investigate what local firms charge for the service. Ask local taxpayers who have received the EITC and purchased a refund product in the past if they'd be willing to share their documentation with you. Often these products are referred to familiarly as "rapid refund" or "fast cash" products. Typically, tax preparers and their bank partners charge a fee for the RAL that is based on the size of the anticipated refund, plus additional flat "documentation" or "loan preparation" fees. With information on the average price for these products in your community, and the number of RALs and RACs requested by EITC filers in years past, you can estimate the amount that low-income filers in your community spend on these high-cost refund products.

[\[Return to top\]](#)

## **HOW MANY PEOPLE IN MY COMMUNITY ARE ELIGIBLE FOR THE EITC, BUT DON'T RECEIVE IT? WHAT'S THE AMOUNT THAT IS "LEFT ON THE TABLE" AS A RESULT?**

The best available research by the IRS and other scholars suggests that between 80 and 85 percent of tax filers who are eligible for the EITC claim the credit. That participation rate exceeds rates for other well-known income support programs like Food Stamps and TANF cash assistance. However, because the EITC can provide a family with such a significant cash infusion, and because a broader range of working families are eligible for the credit than for other means-tested programs, local organizations are devoting significant effort to alerting potentially eligible families about how to claim the EITC.

The IRS continues to work on research methods that will provide better estimates on participation rates in the EITC at a small-area level. (For a detailed explanation of the issues involved in arriving at such estimates, see: [Earned Income Credit Participation—What We \(Don't\) Know](#)) In the interim, because the actual population that claims the EITC changes so much from year to year (one-third of filers who claimed the credit in any given year did not claim it the prior year), and because some families will inevitably miss out on the credit despite the best outreach efforts, the total number of eligible families not claiming the credit in your community may not provide the most useful benchmark for judging outreach efforts.

A more useful approach might be to ask, if you were able to increase the number of eligible filers in your area who claim the credit by 5 percent, how many additional workers and families would benefit, and how many additional EITC dollars would flow into your community? The participation gap is likely to be larger in communities that have more: very low-income working families (incomes under \$10,000); low-income Hispanic families and families whose first language is not English; and families with more than two children. Eligible members of these groups have been found to claim the credit at lower rates than the national average. If your community has large numbers of these types of families, you may have the opportunity to raise the number of families claiming the credit by perhaps as much as 10 percent. To calculate the potential additional number of eligible filers, simply multiply the **eic** value for the most recent year by 5 percent to 10 percent.

Number of EITC-eligible non-filers that an effective outreach campaign could encourage to file = **eic x 5% to 10%**

Research also suggests that eligible filers who fail to claim the credit are typically eligible for somewhat smaller credits on average than those filers who do claim the credit. To calculate the potential economic benefit for these families and your community, multiply the average EITC in your community (**eicam** divided by **eic**) by 50 percent, and multiply the result by the estimated number of additional eligible filers you hope to reach:

EITC dollars that an outreach campaign could add to families and the community = **eic x 5% to 10% x (eicam /eic) x 50% = eicam x 2.5% to 5%**

Finally, the introduction of the Additional Child Tax Credit (ACTC)—the refundable version of the Child Tax Credit (CTC)—increases the refund amounts available to many EITC-eligible filers. National figures show that taxpayers claimed nearly \$26 billion in ACTC in Tax Year 2013, with over 75 percent of those dollars (\$20 billion) going to EITC filers. Users can calculate the relative contribution of the ACTC to taxpayers' refunds by comparing the **actcam** and **refam** variables in the selected geography.

Note that these estimates would provide you with the potential impact on a one-year basis, and would not take into account broader factors that also influence the number of filers in a particular community who receive the credit, including: population growth/decline; changing employment and wage levels; and increases and decreases in the number of children living at home. At the same time, because they represent only one-year estimates, longer-range plans should take into account that these potential economic benefits would recur on an annual basis.

[\[Return to top\]](#)

## **CAN I USE THESE DATA TO CHARACTERIZE THE TOTAL ECONOMIC IMPACT THAT EITC DOLLARS HAVE IN MY LOCAL ECONOMY?**

Economists often describe the total economic impact of a fiscal injection such as the EITC into a local economy through the use of an "economic multiplier." The multiplier represents the factor by which total economic output resulting from the initial investment exceeds that investment, due to the additional economic activity it spurs. The multiplier in any given local economy depends on the interdependence of its different sectors, so it may vary widely from community to community. One example of research on the total economic impact of EITC dollars, for the city of [San Antonio](#), found that refunded EITC dollars spent in the local economy would generate a total economic impact 58 percent larger than those initial expenditures.

[\[Return to top\]](#)