United States Government Accountability Office

Report to the Chairman, Committee on Finance, U.S. Senate

April 2024

TAX ENFORCEMENT

IRS Audit Selection Processes for Returns Claiming Refundable Credits Could Better Address Equity

GAO-24-106126

GAO Highlights

Highlights of GAO-24-106126, a report to the Chairman, Committee on Finance, U.S. Senate

April 2024

TAX ENFORCEMENT

IRS Audit Selection Processes for Returns Claiming Refundable Credits Could Better Address Equity

Why GAO Did This Study

Policymakers and the public expect IRS to administer the tax code fairly. Even though IRS does not collect information on the race and ethnicity of taxpayers, some audit selection criteria and methods could have different implications for taxpayers depending on their race or ethnicity. According to one academic study, audits of EITC returns accounted for 78 percent of the overall estimated racial disparity in audit rates.

GAO was asked to examine IRS's safeguards for ensuring audits do not target filers based on demographic characteristics. This report focuses on audits that address refundable credits, including EITC, and describes IRS policies to address equity in audits; describes relevant audit planning and selection processes; and assesses how IRS considers equity both in developing its audit workplans and in its automated audit selection processes.

GAO reviewed relevant academic literature, analyzed IRS audit tools and procedures, and interviewed IRS officials. GAO assessed IRS audit selection procedures against relevant agency goals and objectives, key practices for evidence-based policymaking, and standards for internal control.

What GAO Recommends

GAO is making six recommendations to IRS, including to calculate the no-change rate without default audits, improve its reviews of audit selection processes, and use additional performance measures in assessing its selection systems. IRS agreed to all of GAO's recommendations.

View GAO-24-106126. For more information, contact Jessica Lucas-Judy at 202-512-6806 or lucasjudyj@gao.gov.

What GAO Found

The Internal Revenue Service (IRS) and the Department of the Treasury have committed to tax administration equity, including equity in how IRS conducts audits, in multiple policies and various media. IRS also has begun to examine potential systemic biases in audit enforcement across several demographic characteristics, including age, gender, geography, race, and ethnicity.

IRS's Wage & Investment Division (W&I) audits returns claiming refundable tax credits, including the Earned Income Tax Credit (EITC) and the Premium Tax Credit. W&I uses a mix of manual and automatic processes to determine the number and types of audits it will conduct and relies on four objectives to develop its audit workplan. One of these objectives is the no-change rate—the percentage of returns that will yield no additional revenue after audit.

IRS seeks to achieve a low no-change rate because that indicates it is auditing noncompliant taxpayers. However, the calculation of the no-change rate includes default audits—audits closed as a "change" because taxpayers did not respond or provided insufficient responses to IRS's notices. IRS officials said their recent research found that Black taxpayers are more likely not to respond to IRS correspondence than taxpayers of other races. Default audits also may be more common among low-income and EITC taxpayers, because of challenges that make communicating successfully with IRS more difficult, such as being transitory or not having bank accounts. W&I uses past results to inform current audit planning, which could lead to W&I disproportionately selecting the types of returns that have historically resulted in filers' nonresponse rather than in confirmed noncompliance.

The primary system IRS uses to select specific returns for audit is the Dependent Database (DDB) program, an automated system that flags returns for potential risk of noncompliance. While IRS regularly reviews the program, the review process does not comprehensively consider data inputs and assumptions that could inform IRS about the demographic equity of the audit selection process, creating the potential for unintended bias in audit selection. For example, GAO found that some risk scores contained in the DDB program vary by sex, which could skew selection, and have not been updated since 2001.

The primary measure for assessing the performance of the DDB system is the no-change rate. IRS officials said they are piloting a new scoring model to better detect noncompliance in tax return filings, which would be used for audit selection. IRS used the amount of revenue gained through conducting the audits as a performance measure for the pilot, an indicator of the extent of noncompliance rather than just its presence. The agency may be missing opportunities to improve the likelihood that IRS is properly identifying returns at highest risk of noncompliance if it does not consider additional performance measures in reviewing its automated audit selection process.

Contents

Abbreviations

AI	Artificial Intelligence
DDB	Dependent Database
EITC	Earned Income Tax Credit
EPST	Enterprise Planning Scenario Tool
FCR	Federal Case Registry
IRM	Internal Revenue Manual
IRS	Internal Revenue Service
RIVO	Return Integrity Verification Operations
SRA	Systems Research and Application
TIGTA	Treasury Inspector General for Tax Administration
W&I	Wage and Investment Division

U.S. GOVERNMENT ACCOUNTABILITY OFFICE
441 G St. N.W.
Washington, DC 20548

April 25, 2024

The Honorable Ron Wyden
Chairman
Committee on Finance
United States Senate

Dear Mr. Chairman:

Policymakers and the public expect the Internal Revenue Service (IRS) to administer the tax code fairly. IRS has stated its commitment to this goal in congressional hearings, strategic plans, and other publications. Administering a tax system involves multiple tasks, including choosing which returns to audit and conducting and closing audit investigations. As IRS seeks to administer the tax code fairly, it also has other commitments, such as minimizing the tax gap and ensuring audit coverage across various types of returns.

IRS generally aims to audit returns that are most likely to be noncompliant, whether intentionally or by mistake. Audit selection criteria and methods, including those established in automated processes, play an important role in promoting fairness. It is critical to understand the different implications these selection methods could have for taxpayers depending on their race, ethnicity, income level, or other demographic and economic characteristics.

Recently, members of Congress, academics, and the Department of the Treasury have raised concerns about possible inequity in tax administration, including in audit rates. For instance, in a letter to IRS Commissioner Danny Werfel, several members of the U.S. House of Representatives requested a detailed plan and timeline for IRS to implement corrective actions to address racial disparities in audit selection.[1] Furthermore, academics and government researchers have written about how audit rates vary for individual taxpayers of different

[1]Committee on Ways and Means, Subcommittee on Oversight, U.S. House of Representatives, Letter to Commissioner Werfel (Mar. 9, 2023). The letter was submitted by the Ranking Member and other members of the subcommittee.

races.[2] The President has also raised broader concerns about equity and fairness in government, and he announced "a comprehensive government approach to advancing equity for all."[3] He defined equity as "the consistent and systematic treatment of all individuals in a fair, just, and impartial manner, including individuals who belong to communities that often have been denied such treatment."[4]

In September 2023, IRS announced an effort to restore fairness to the tax system by shifting more attention onto high-income earners, partnerships, large corporations, and promoters abusing the nation's tax laws.[5] This effort will include a focus on making improvements in audits involving the Earned Income Tax Credit (EITC) to advance equity in audits, according to IRS.

We have previously reported on trends in IRS audit rates and results for individual taxpayers by income.[6] Our work showed that the audit rate for returns claiming the EITC is higher than the average audit rate. In another report, we estimated that there were disparities in the amount of certain tax provisions, including the EITC, by the race and ethnicity of taxpayers.[7]

You asked us to examine IRS's safeguards for ensuring that audits do not target filers based on demographic characteristics. Additionally, the IRA included a provision for GAO to support oversight of the use of funds appropriated in the IRA, including ensuring the impacts of funding decisions are equitable.[8] This report focuses on audits of tax returns

[2]Hertz, T. et al., "Differences in Audit Rates by Race," Presented at the 13th Annual IRS/TPC Joint Research Conference on Tax Administration, June 22, 2023. Elzayn, H. et al., *Measuring and Mitigating Racial Disparities in Tax Audits,* Stanford Institute for Economic Policy Research (Jan. 30, 2023). These studies use statistical methods to predict the race of the filer to analyze differences in audit rates.

[3]Exec. Order No. 13985, 86 Fed. Reg. 7009 (Jan. 25, 2021).

[4]Exec. Order No.14091, 88 Fed. Reg. 10825, 10831-32 (Feb. 22, 2023).

[5]Promoters are individuals who arrange and market abusive tax schemes designed to circumvent tax laws or evade taxes.

[6]GAO, *Tax Compliance: Trends of IRS Audit Rates and Results for Individual Taxpayers by Income*, GAO-22-104960 (Washington, D.C.: May 17, 2022).

[7]GAO, *Tax Equity: Lack of Data Limits Ability to Analyze Effects of Tax Policies on Households by Demographic Characteristics*, GAO-22-104553 (Washington, D.C.: May 18, 2022).

[8]Pub. L. No. 117-169, § 70004, 136 Stat. 2087 (2022).

claiming refundable credits, including the EITC, and (1) describes IRS policies to address equity in audits; (2) describes relevant IRS audit planning and selection processes; (3) assesses equity considerations in IRS processes to develop audit workplans; and (4) analyzes how IRS addresses equity in its automated audit selection processes.

To address each of our objectives, we interviewed IRS officials with knowledge about the tools and procedures used to select returns for audit. We also interviewed relevant IRS officials concerning their understanding of IRS's Wage and Investment Division's (W&I) mission, objectives, and internal controls related to W&I's audit selection.

To address our first objective, we reviewed agency strategic plans along with other public documents and statements provided by IRS and Treasury. We analyzed this information to identify IRS's audit equity considerations.

For our second objective, we analyzed W&I's audit tools and procedures. W&I is responsible for audits that address refundable credits such as the EITC and the Additional Child Tax Credit. This report focuses only on W&I audit tools and procedures to identify taxpayers who are potentially noncompliant because these procedures are the primary method used to select returns that claim the EITC and other refundable credits for audit. Previous academic research suggests that IRS's audit treatment of returns claiming refundable credits might explain audit inequity by demographic characteristics.

To address our third objective, we reviewed the guidance and processes IRS uses to develop audit workplans. To address our fourth objective, we reviewed IRS's automated audit procedures, including algorithms used as part of the process to select returns to audit. We reviewed these procedures to analyze the methods and data used to select returns.

We assessed IRS's consideration of equity in its audit selection procedures in our third objective and artificial intelligence (AI) audit tools in our fourth objective, respectively, against relevant agency goals and objectives, our past work on key practices for evidence-based policymaking, and monitoring AI systems.[9] Additionally, to address our

[9]GAO, *Evidence-Based Policymaking: Practices to Help Manage and Assess the Results of Federal Efforts*, GAO-23-105460 (Washington, D.C.: July 2023); and *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities*, GAO-21-519SP (Washington, D.C.: June 2021).

fourth objective, we assessed IRS audit selection procedures against internal control standards to determine the extent to which they (1) appropriately included control activities to achieve objectives and respond to risks of inequity in audits and (2) used quality information to mitigate inequity in audits. Those principles of internal control are outlined in *Standards for Internal Control in the Federal Government*.[10] See appendix I for additional information on our scope and methodology.

We conducted this performance audit from June 2022 to April 2024 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Background

Recent research indicates that there may be disparities in IRS audits by taxpayer demographics, such as race, sex, family composition, filing status, location, and income.

One article in 2019 estimated disparities in EITC audits by geographic location. It found that, similar to the distribution of returns claiming the EITC, audit rates were generally more concentrated in the southeastern U.S. than in several other regions.[11] Another article, which cited and expanded on the first, noted that the estimated geographic concentration of audited returns includes areas that have higher populations of Black and Hispanic taxpayers.[12] W&I officials said their division's selection of returns to audit does not use geographic information.

In 2021, the Treasury Inspector General for Tax Administration (TIGTA) found differences in the percentage of EITC returns that IRS selected for audit by state.[13] According to TIGTA, states with the highest percentage of EITC claims that were examined did not always rank as the states with the highest percentage of overall claims filed. However, TIGTA also noted

[10]GAO, *Standards for Internal Control in the Federal Government*, GAO-14-704G (Washington, D.C.: September 2014).

[11]Bloomquist, Kim, "Regional Bias in IRS Audit Selection," *Tax Notes* (Mar. 4, 2019).

[12]Kiel, Paul and Hannah Fresques, "Where in The U.S. Are You Most Likely to Be Audited by the IRS?", *ProPublica* (Apr. 1, 2019).

[13]Treasury Inspector General for Tax Administration, *The Earned Income Tax Credit Examination Compliance Strategy Can Be Improved*, 2021-30-051 (Sept. 2, 2021).

that IRS's audit selection process focuses on areas with the highest likelihood of noncompliance and that IRS's filters found more indications of likely noncompliance in the states with the highest percentage of EITC claims that were examined.

IRS officials noted that their data indicate geographic variation in the audit rate of EITC returns that are flagged by IRS filters as being at the highest risk of noncompliance. It is unclear what is driving these differences in the share of potentially noncompliant returns that are audited. One possible driver could be variance in IRS's final assessment of the risk of returns' noncompliance.

As noted earlier, we previously reported that IRS audited taxpayers claiming the EITC at a higher rate than the average rate for all taxpayers.[14] IRS officials explained that the EITC audit rates can be higher than audit rates for some higher-income taxpayers because EITC audits are limited in scope and less time consuming, allowing IRS to conduct more audits. They also said that returns claiming the EITC have high rates of improper payments and require a greater enforcement presence in the form of audits.[15] Overall, as we reported, IRS generally audited higher-income taxpayers at higher rates than lower-income taxpayers from tax years 2010 through 2019.

More recently, a study conducted by authors from Stanford University, IRS, and other affiliations identified estimated disparities in audits by race.[16] To impute the race of taxpayers, the authors used known information, such as name and geographic location. They estimated that Black taxpayers are audited at higher rates than taxpayers of other races. This disparity in audit rates between Black EITC claimants and EITC claimants of other races is substantially larger than the disparity between Black taxpayers and taxpayers of other races within the full population, according to the study's estimates. The authors also estimated

[14]GAO-22-104960.

[15]The Payment Integrity Information Act of 2019 increased federal agencies' requirements to report improper payments. Pub. L. No. 116-117, 134 Stat. 113 (Mar. 2, 2020). According to IRS officials, IRS must develop and submit a corrective action plan to the Office of Management and Budget annually that includes improper payment root cause identification, reduction targets, and identification of accountable officials. Audits, including audits of returns claiming the EITC, have been a part of the corrective action plan.

[16]Elzayn, H. et al., *Measuring and Mitigating Racial Disparities in Tax Audits*.

significantly smaller audit disparities by race among taxpayers who do not claim the EITC.

Additionally, the authors found that the racial audit disparity within EITC returns contributes an estimated 78 percent of the overall disparity.[17] For example, among unmarried EITC claimants with dependents, the audit rate for Black men is more than 4 percentage points larger than the audit rate for men who are not Black.[18] The authors observed smaller audit disparities by race among joint filers, unmarried women, and unmarried men who do not claim dependents.

IRS does not collect data about the race and ethnicity of taxpayers. Consequently, when IRS predicts each return's risk of noncompliance with the tax code, the agency cannot directly identify the race or ethnicity of filers it deems most at risk of noncompliance. However, current IRS research efforts have also confirmed the existence of racial disparities in audits and have identified unintentional algorithmic biases as a potential source of disparities. Specifically, that research noted (1) limitations in the data used to determine residency and relationship tests for EITC eligibility, and (2) outdated models as possible contributions to algorithmic bias and, consequently, racial disparities in audits.[19]

While IRS research has confirmed that IRS's audit selection algorithms could be possible drivers of unintentional biases, variation in compliance rates across different demographic groups could also affect audit outcomes. Researchers have suggested several potential causes could explain the potential lower levels of compliance by certain populations.[20] For example, those populations might disproportionately use complex provisions that are harder to claim properly, such as the EITC. Alternatively, on average, they could have limited access to quality tax

[17]IRS subject matter experts noted that estimates of disparities in audit rates are preliminary and may change based on ongoing research. For example, IRS has begun research to test and refine race imputation approaches, and changes to the imputation methodology may yield different conclusions about the extent of disparity in audits.

[18]Compared to the overall audit rate, among unmarried EITC claimants with dependents, the audit rate for Black men is approximately 14 times higher. For men who are not Black, the audit rate is approximately six times higher.

[19]Hertz, T. et al., "Differences in Audit Rates by Race."

[20]National Taxpayer Advocate, "Annual Report to Congress," 2007.

preparers and the internet. Language barriers could also limit their engagement with IRS.

IRS Leadership Expressed Commitment to Equity and Fairness in Audits

IRS and Treasury have committed to tax administration equity, including equity in how IRS conducts audits, in multiple policies and various media. In both its *IRS Strategic Plan FY 2022-2026* and its *Internal Revenue Service Inflation Reduction Act Strategic Operating Plan FY 2023-2031*, IRS commits to administering the tax code equitably. The agency states that it will enforce the tax code with integrity, transparency, objectivity, and fairness, and IRS will use advanced technologies to help it accomplish this goal. For example, the plans state that machine learning will improve IRS's understanding of customer behavior, which will aid the agency in selecting returns for audit.

Furthermore, IRS has begun to examine potential systemic biases in audit enforcement across several demographic characteristics, including age, gender, geography, race, and ethnicity. In January 2023, IRS began internal research dedicated to understanding recent academic work that estimated disparities in audit rates by demographic characteristics. IRS's work is ongoing, but the agency presented interim findings publicly in June 2023.[21]

Other IRS documents and policies, such as the Internal Revenue Manual (IRM), commit the agency to fairness in tax administration. The IRM prohibits discrimination on the basis of race, color, national origin, sex, religion, disability, and age.[22] In a website post in May 2023, the agency stated that enforcement of the tax law is a key component in promoting the highest degree of voluntary compliance, and IRS concluded that it is being fair to those who are compliant by pursuing filers who are not.[23]

IRS also announced in September 2023 that it would be increasing its focus on the wealthy, partnerships, and other high earners. According to its announcement, there have been sharp drops in audit rates for these taxpayer segments during the past decade. IRS also stated, "as part of the effort, the IRS will also ensure audit rates do not increase for those

[21]Hertz, T. et al., "Differences in Audit Rates by Race."

[22]IRM § 1.20.4.2 (08-12-2020).

[23]IRS, "IRS Update on Audits" (Washington, D.C.: May 1, 2023), accessed December. 6, 2023, https://www.irs.gov/newsroom/irs-update-on-audits.

earning less than $400,000 a year as well as adding new fairness safeguards for those claiming the Earned Income Tax Credit."[24]

IRS Commissioners have made public statements about fairness and equity in tax enforcement. During a House Ways and Means Committee hearing on accountability in IRS, Commissioner Werfel said that he believes fairness and equity are essential components in the tax system.[25] He committed to studying whether the enforcement of the EITC was inconsistent with these values. Former IRS Commissioner Charles Rettig, in a statement about his tenure, also committed to the principle of fairness in tax administration through appropriate and meaningful enforcement efforts.[26]

On June 8, 2023, Treasury released a memorandum for department bureaus and offices entitled, *Principles for Promoting Fair and Effective Compliance*.[27] The memorandum includes core principles and guidelines to help ensure fairness in enforcement efforts. One principle is to promote fairness, trust, and accountability in enforcement efforts. The memorandum notes that all bureaus should regularly audit their systems to determine whether compliance and enforcement practices result in unintended and unwarranted disparities across race, ethnicity, gender, age, marital status, geography, or other factors. When disparities are identified, bureaus should employ corrective action as swiftly as possible.

IRS Uses Manual and Automatic Processes to Identify Returns for Audit

IRS's Wage and Investment Division (W&I) uses various methods to identify taxpayers who are potentially noncompliant with the tax code. The majority of W&I's examinations are conducted by the Refundable Credits Examination Operations office. While most refundable credit returns audited by this office claim the EITC, it also examines returns claiming other refundable credits on the Form 1040 U.S. Individual Income Tax Return, such as the Premium Tax Credit. When tax credits are

[24]IRS, "IRS announces sweeping effort to restore fairness to tax system with Inflation Reduction Act funding; new compliance efforts focused on increasing scrutiny on high-income, partnerships, corporations and promoters abusing tax rules on the books," IR-2023-166 (Sept. 8, 2023).

[25]House Ways and Means Committee, Hearing on Accountability and Transparency at the Internal Revenue Service (Washington, D.C.: Apr. 27, 2023).

[26]IRS, "IRS Statement—IRS Commissioner Charles Rettig's Final Message" (Nov. 10, 2022).

[27]Department of the Treasury, Memorandum for Department Bureaus and Offices, *Principles for Promoting Fair and Effective Compliance* (Washington, D.C.: June 8, 2023).

refundable, the amount claimed is payable to the taxpayer as a refund to the extent that the tax credit exceeds the taxpayer's tax liability. Table 1 provides data on the share of audits conducted by the office by return type.

Table 1: Share of Planned Refundable Credit Return Audits by Return Type, Fiscal Year 2023

Return type	Percentage
Earned Income Tax Credit (EITC)	78.1
Non-EITC	21.9
Other tax credits	7.9
Premium Tax Credit	6.6
Non-EITC audit program returns	5.1
Family tax credits	2.3

Source: GAO analysis of Refundable Credits Examination Operations Workplan. | GAO-24-106126

Note: Family tax credits include the Additional Child Tax Credit, Refundable Child Tax Credit, Refundable Child and Dependent Care Credit, and Sick and Family Leave Credit. Other tax credits include American Opportunity Tax Credit, Health Coverage Tax Credit, and Fuel Tax Credit. Non-EITC Audit Program Returns include returns selected under specific audit program issues, such as the Questionable Refund Program.

W&I audits returns both prior to and after filers receive refunds, called pre-refund and post-refund audits, respectively. All W&I pre- and post-refund compliance efforts occur via correspondence—audits conducted through the mail. According to IRS officials, W&I conducts most audits pre-refund, thereby freezing the refundable credit portions of taxpayers' refunds until the audits conclude.

Officials from W&I said the division calculates its monthly workload prior to selecting returns to audit. That process combines manual and automated methods to determine the final audit workplan. Specifically, W&I uses the Enterprise Planning Scenario Tool (EPST) model to determine the number and types of audits it will conduct monthly. This tool is a quantitative, predictive model based on historical audit examination results, such as the no-change, agreement, default, and response rates.

W&I staff adjust the EPST's inputs and receive outputs that describe the volume of returns to be audited by case type (defined by credits, or combinations of credits, claimed on tax returns), the month cases will be audited, and the IRS campus where the work will be done. Staff minimize or maximize value in four adjustable objectives, which are the EPST inputs:

- dollars protected and assessed, or the amount of revenue gained through conducting an audit;
- no-change rate, or the percentage of returns that will yield no additional revenue after audit;
- cycle time, or how long it takes to audit a return; and
- response rate, or how many audited taxpayers will communicate with IRS via correspondence.

These objectives are also subject to several constraints, such as available staff hours and available inventory of tax returns with indicators of noncompliance. EPST users develop a variety of scenarios based on adjusted objectives and resource constraints. According to IRS officials, in recent years, maximizing dollars protected and assessed has been chosen as the criterion used to develop workplans. Management reviews staff allocated for that fiscal year and other factors to determine the final workplan.

Next, W&I selects individual returns using both manual and automated processes, including statistical models. Most W&I audits are selected through automated processes. According to W&I officials, audit selection operates under a comprehensive set of checks and balances aimed at delivering a fair and equitable process to find taxpayers most likely to be noncompliant. These processes include the following:

- The **Dependent Database** (DDB) is an automated system W&I uses to identify potential noncompliance with refundable credits on individual tax returns. Most W&I audits are selected through the DDB.
- **Return Integrity Verification Operations** (RIVO) refers returns to W&I that are detected as potentially noncompliant in the RIVO computer program. The majority of the referrals to W&I Exam from RIVO are based on the Return Review Program to identify overreported income and withholding discrepancies that would impact refundable credits claimed on the Form 1040 series returns.
- **Unallowables** are items reported on tax returns that are not allowed by law and result in IRS issuing a hold which prevents the amount of the unallowable from being refunded or offsetting tax liability. Unallowables can include taxpayers' claims of deductions, items

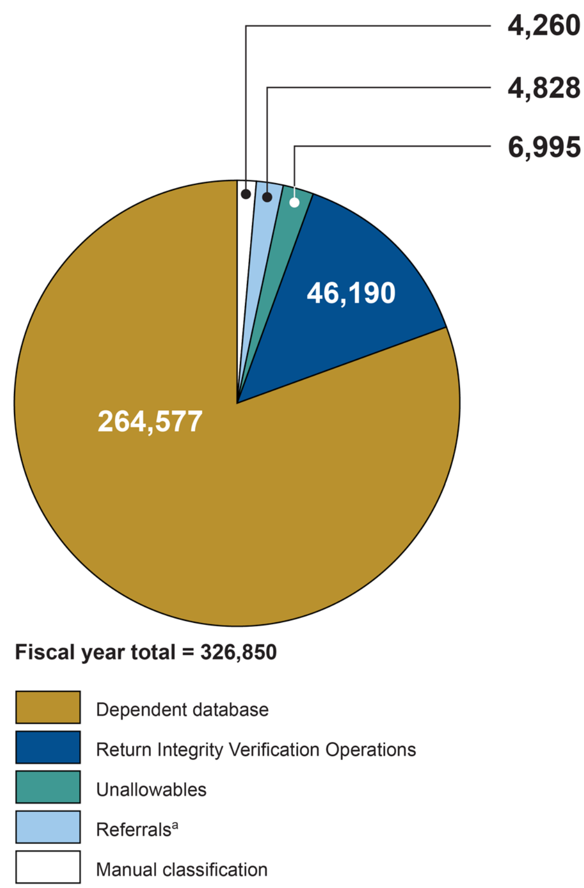
exceeding statutory limits, or items not supported by the proper information or schedule.

- A **referral** is a return sent to W&I's Refundable Credits Examination Operations office for audit consideration from another IRS office or an outside entity, such as state agencies.[28]
- **Manual classification** is the process in which an audit staff member determines whether a return should be selected for audit and what issues IRS should audit.

Figure 1 shows the number of audits conducted by W&I by audit selection method for fiscal year 2022.

[28]We have previously reported on IRS's use of outside referrals. GAO, *Abusive Tax Schemes: Additional Steps Could Further IRS Efforts to Detect and Deter Promoters*, GAO-23-105843 (Washington, D.C.: Dec. 15, 2022), discusses how outside referrals help IRS identify abusive tax schemes. GAO, *Priority Open Recommendations: Internal Revenue Service*, GAO-23-106470 (Washington, D.C.: July 31, 2023), includes our prior recommendations for improving the outside referral process.

Figure 1: Number of Wage and Investment Division Audits by Selection Method, Fiscal Year 2022



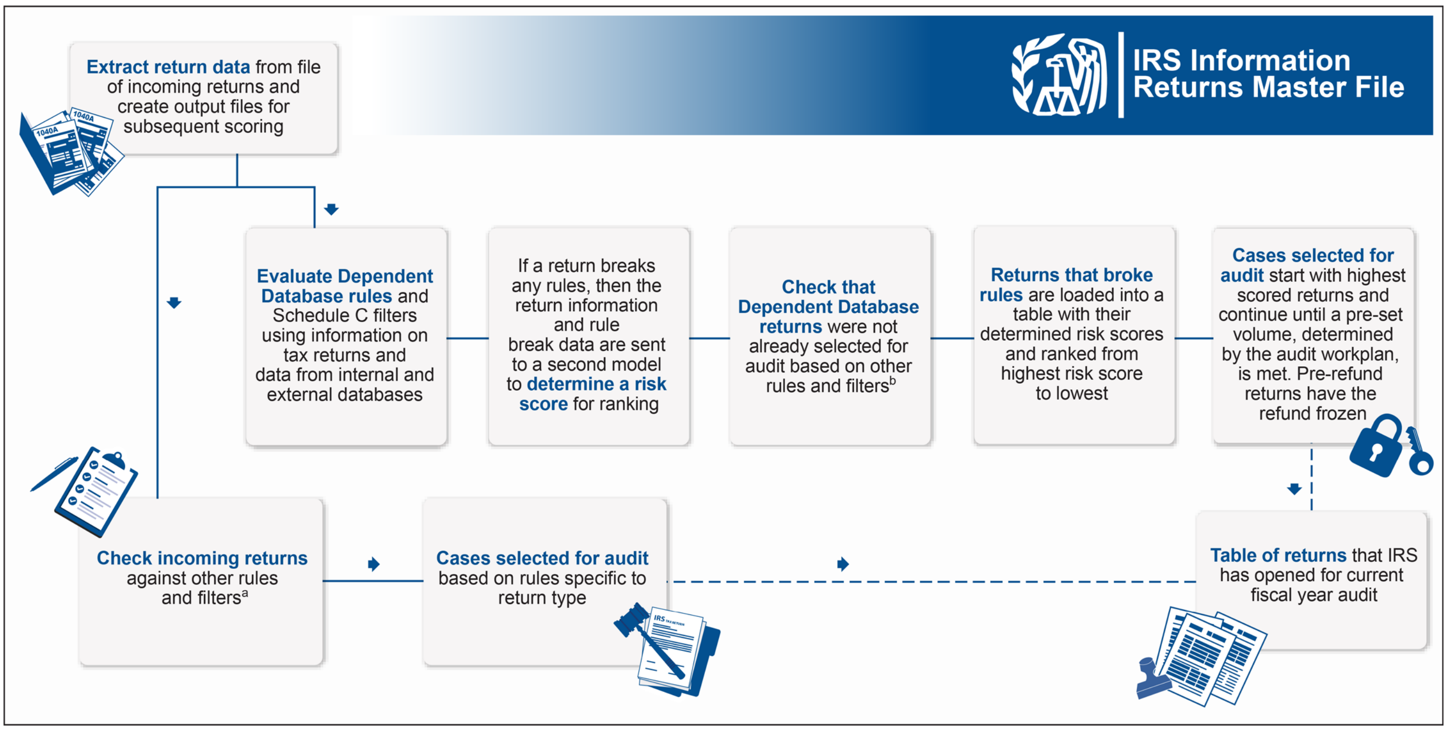
Source: GAO analysis of IRS information. | GAO-24-106126

[a]Referrals include outside referrals. The value of outside referrals was too small to report separately and does not reflect normal levels. According to Internal Revenue Service officials, they would expect this value to be between 1,000 and 1,500 more.

The DDB system identifies potentially noncompliant returns during processing, using data from tax returns and external data matching processes. It relies on two key external databases: (1) child support orders and custodial status in the Federal Case Registry from the Department of Health and Human Services; and (2) Kidlink, the Social Security Administration's database of Social Security numbers of parents linked with newborn children.

The DDB selection process uses these external databases and a series of rules applied to tax returns to identify potentially noncompliant returns. A risk-based scoring model then assigns a score to those returns based on a series of algorithms. Figure 2 demonstrates the DDB audit selection process.

Figure 2: Dependent Database Audit Selection Process



Source: GAO analysis of IRS data. | GAO-24-106126

[a]In addition to Dependent Database rules and filters, the system also includes other filters to select specific types of returns based on other issues related to refundable credits.

[b]Returns are checked against those that were chosen for other reasons and filters and are removed prior to ranked selection.

According to IRS officials, duties are separated throughout the audit selection process. The employees who examine selected returns are not involved in developing DDB selection rules, planning decisions, or any other rule-based programming.

After W&I completes an audit, the result is either a recommended change to the tax return or the taxpayer sufficiently supports information on the tax return, which results in no change to the initial filing. According to IRS and W&I officials, a low no-change rate indicates that the audit processes effectively found noncompliant taxpayers and, therefore, are effective at meeting one measure of fairness.

W&I also closes some audits because of taxpayer nonresponse. These are termed default audits. Default audits also include audits closed because the taxpayer response was insufficient to support information provided on the tax return. In the case of pre-refund audits—audits conducted prior to IRS sending refunds to filers—default audits result in an automatic change to the returns, with the refunds being withheld. According to IRS officials, the share of DDB audits that were closed due to taxpayer nonresponse ranged from 25 percent to 36 percent between fiscal years 2017 and 2022. Among default audits in fiscal year 2023, about 60 percent were from taxpayer nonresponse.

IRS Process for Determining Its Audit Workplan Does Not Adequately Address Equity

A Key IRS Audit Selection Target May Not Adequately Reflect Equity

According to W&I officials and IRS audit policy, IRS attempts to select returns to audit that are not fully compliant, which avoids burdening compliant taxpayers. This means pursuing a low no-change rate. This approach is intended to reduce taxpayer burden and provide a higher return on investment. IRS policy documents, such as the Internal Revenue Manual (IRM), also consider auditing noncompliant taxpayers and not burdening those who are compliant as a measure of fairness. The no-change rate reflects that consideration by measuring the share of audits that did not result in a change to the tax return—that is, the share of the audited returns that were determined to be compliant.

However, according to IRS officials, the calculation of the no-change rate includes pre-refund default audits. They also said default audits include audits closed because the taxpayer did not respond to IRS notices, as well as many cases in which taxpayers responded but did not provide sufficient documentation to support their eligibility. Default audits also include cases in which taxpayers never sign the final exam report agreeing to changes. W&I designates pre-refund default audits as a "change" and does not differentiate them from audits that resulted in a change and in which the taxpayer responded.

It is possible that some default audits would have concluded with no change in tax liability if a taxpayer had responded to IRS communications. We have previously reported on the difficulties of interpreting a low no-change rate when it includes default audits.[29] IRS has gathered information to try to understand why taxpayers might not respond to correspondence from the agency. According to IRS officials, it is difficult to study taxpayer nonresponse by virtue of these taxpayers' lack of engagement with IRS. IRS officials said they continue to try to reach nonrespondent taxpayers to encourage them to participate in the audit process. IRS officials acknowledged there are likely eligible taxpayers among the nonrespondents, which could result in estimates of the noncompliance rate that are skewed higher.

IRS officials stated their recent research into this issue analyzed differences in taxpayer response based on demographics, finding that Black taxpayers are more likely not to respond to IRS correspondence than taxpayers of other races. In addition, according to a National Taxpayer Advocate's annual report, default audits may be more common among low-income and EITC taxpayers. Based on data cited in that report, taxpayers with low incomes are more likely to experience lower literacy rates and often possess limited English proficiency. Those data also indicated that it is more likely these taxpayers do not have bank accounts, which can affect their ability to substantiate income and expenses. Lastly, the report notes that these taxpayers also tend to be more transitory, a factor that hurts their ability to receive IRS correspondence and respond to it in a timely manner.[30]

According to IRS officials, past research that addressed, to some degree, the probable compliance of audit nonrespondents did not indicate that a large number of nonrespondents would be eligible for the EITC. IRS officials also stated that in their recent research on taxpayer nonresponse, they estimated noncompliance rates are overstated to a similar degree for both Black taxpayers and taxpayers of other races.

[29]GAO, *Refundable Tax Credits: Comprehensive Compliance Strategy and Expanded Use of Data Could Strengthen IRS's Efforts to Address Noncompliance*, GAO-16-475 (Washington, D.C.: May 27, 2016); and *IRS Correspondence Audits: Better Management Could Improve Tax Compliance and Reduce Taxpayer Burden,* GAO-14-479 (Washington, D.C.: June 5, 2014).

[30]National Taxpayer Advocate, "Annual Report to Congress" (2021).

They also indicated plans to explore these issues further in their pilot models with and without data from nonresponse audits.

W&I's decision to include default audits in the no-change calculation could misleadingly reduce the no-change rate and have implications for the equity of tax audits. W&I uses past results from audit outcomes to inform current audit selection planning. For example, W&I considers historical audit results of different case types, and it assumes that the results of future audits will be similar to those of past audits. By relying on a no-change rate that is likely lowered by the inclusion of audits based on taxpayer nonresponse, W&I could be disproportionately selecting returns that are more likely to be default audits rather than those likely to be confirmed as noncompliant. Further, in pursuit of a low no-change rate, W&I could be inadvertently selecting taxpayers with certain demographic characteristics disproportionately.

We have previously identified key practices for evidence-based policymaking and performance management. One key practice is for agencies to assess the extent to which existing evidence meets organizational needs for learning and decision-making.[31] In doing so, agencies should also ensure new evidence meets relevant quality standards. Those standards include federal government-wide and organization-specific legal requirements, and related policies, guidance, and leading practices. The no-change rate represents an important source of evidence used in deciding on the audit workplan.

Calculating the audit no-change rate both with and without default audits included could help IRS understand the effect these audits have on the no-change rate and subsequently the audit workplan. It could also distinguish between default audits based on taxpayer nonresponse and those based on taxpayer response that is insufficient to support information provided on the tax return. IRS officials said they are considering further work on their audit selection models to determine the effect of nonresponse on demographic equity of audits. Detailed analysis of no-change rates would build on IRS's recent research to increase W&I's ability to determine how different audit types intersect with taxpayers' demographic traits.

[31]GAO-23-105460.

IRS Does Not Have Guidance for Considering Its Demographic Equity Research in Audit Workplan Development

IRS seeks to balance several goals in its workplan development. However, W&I has not developed guidance on how to consider demographic equity within the context of developing audit workplans.

W&I officials consider their audit selection process as fair in its attempt to find noncompliant taxpayers. In 2015, we recommended IRS clearly define and document the key term "fairness" for return selection activities to help ensure IRS's Small Business/Self Employed Division's audit selection program meets its mission and selects returns fairly.[32] In response to our recommendation, IRS developed the following three-part definition of fairness in enforcing the tax law: (1) pursuing those who fail to voluntarily comply or otherwise meet their tax obligations, (2) using an equitable process that selects returns for audit based on the likelihood of reporting errors across all areas of potential noncompliance, and (3) respecting and adhering to the rights of taxpayers.[33] While this explanation of fairness partially relies on a concept of equity based on compliance, it does not explicitly consider equity based on demographic characteristics, such as race.

IRS's policies and plans state that IRS intends to conduct enforcement selection processes equitably. For example, IRS's Inflation Reduction Act Strategic Operating Plan calls for the agency to employ centralized, analytics-driven, risk-based methods to aid in the selection of compliance cases while aiming to sustain taxpayer compliance and ensure fairness in case selection. The plan also states that IRS will assess its enforcement actions to ensure fairness across demographic populations, including age, gender, geography, race, and ethnicity.

W&I has detailed guidance for developing audit workplan scenarios using the Enterprise Planning Scenario Tool (EPST). In addition, managers consider a number of factors when determining the final audit workplan. However, W&I does not have guidance on how to consider equity by taxpayers' demographics in the context of developing the audit workplan.

[32]GAO, *IRS Return Selection: Certain Internal Controls for Audits in the Small Business and Self-Employed Division Should Be Strengthened,* GAO-16-103 (Washington, D.C.: Dec. 16, 2015).

[33]IRM § 1.2.1.2.36(3) (10-24-2016).

IRS does not collect data on taxpayers' race, which makes race-conscious equity analyses challenging.[34] According to IRS officials, IRS's Office of Research, Applied Analytics and Statistics is currently using imputation methods to analyze the demographic and economic distribution of default audits and different return types, such as EITC returns.[35] W&I officials stated that predictions about taxpayers' demographic characteristics have not been and will not be used in the operational selection of returns to audit. Still, continuing this work could help the division determine the potential for biases—racial and otherwise—that may be unintentionally embedded in its audit workplans and processes.

W&I officials said they have observed that different optimization weights in EPST audit plans tend to produce similar results. For instance, given the same resource constraints, an audit plan based solely on protecting revenue is similar to an audit plan based on optimizing cycle time. This information indicates that IRS might be able to consider other goals in developing the workplan, such as demographic equity, without compromising other agency goals, such as optimizing revenue.

IRS has stated that based on the equity research it has already conducted, it plans to reduce the number of correspondence audits focused specifically on certain refundable credits, including the EITC, that are included in its audit workplan. However, W&I does not have guidance for how to systematically consider the results of IRS's demographic equity research in developing its audit workplan. Without clear guidance, IRS risks introducing demographic biases introduced into the selection process.

[34]We recently reported on these challenges and identified several methods to conduct this work, including matching IRS data with Census data and the use of imputation methods to predict the race of taxpayers. While it is possible for the Census Bureau and IRS to share data, there are current limitations to implementing an interagency sharing option. Legal protections on data restrict agencies' abilities to systematically share data. These laws require the confidentiality of information collected. For example, Titles 13 and 26 of the United States Code limit the ability of the Bureau and IRS to share data. In GAO-22-104553, we recommended that Congress consider revising the laws to enable better inter-agency data sharing. As of March 2024, Congress had not taken action on this matter.

[35]IRS is using the Bayesian Improved First Name Surname and Geocoding method to predict the race of taxpayers. Government and academic researchers, including ourselves, have used this method or similar methods, which uses name and geolocation to statistically predict a person's race.

Gaps in Oversight of Automated Audit Selection Processes Have Implications for Equity

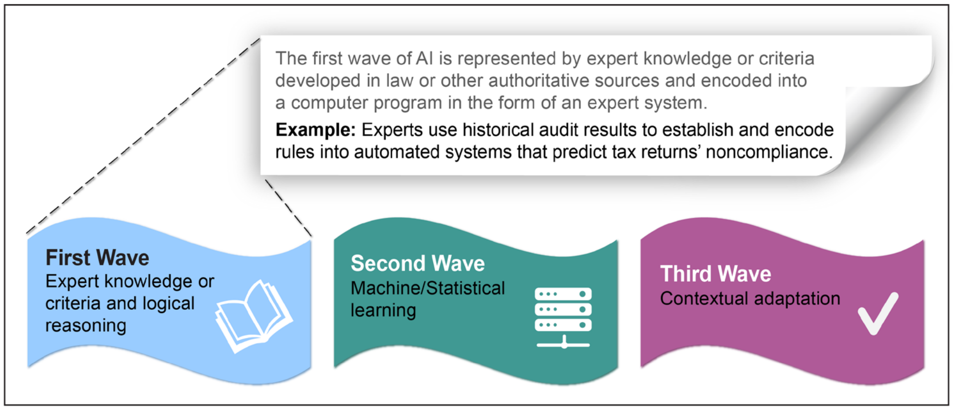
IRS Uses Several Automated Systems to Select Tax Returns for Audit

The primary system W&I uses to select returns claiming refundable credits for audit is the Dependent Database (DDB) program. DDB uses “rules” to flag tax returns for potential risk of noncompliance. The rules include the appearance of duplicate taxpayer identification numbers, filing status, and information on types of income and credit values, among others. The DDB program’s rules also include checks for the accuracy of the dependents listed on tax returns. These checks are done by verifying those dependents using outside databases, specifically the Federal Case Registry (FCR), from the Department of Health and Human Services, and Kidlink, a database from the Social Security Administration.

FCR data may be available for a child if a divorce decree was issued or if the child receives public assistance. Generally, Kidlink contains information about children and their birth parents for children born after 1998. If information regarding dependents claimed on the return is contradicted by these databases, the return is automatically flagged as risky for noncompliance by the DDB program. Conversely, when tax return information appears to be confirmed by these databases, the return is less likely to be flagged as risky.

In addition to the rules contained directly in the DDB, the program also uses another model, the Systems Research and Application (SRA) model, to determine risk scores for returns. Tax returns flagged by the DDB program are then evaluated by the SRA model to determine their risk scores. We consider DDB a first-wave artificial intelligence (AI) system, which means it comprises expert knowledge encoded into a computer system (see fig. 3).

Figure 3: The First Wave of Artificial Intelligence (AI) in Internal Revenue Service (IRS) Audits



Source: GAO analysis of IRS information. | GAO-24-106126

According to IRS officials, SRA is a data mining and machine learning model used to identify patterns in historical audit data to reveal predictors of audit outcomes. Although it is contained within the DDB program, it contains a set of rules separate from DDB to predict the likelihood of "no change" or "change" to returns. When a return triggers a DDB rule, SRA evaluates the risk of that return resulting in a "change" or "no change." Within SRA, each rule is assigned a confidence value for its prediction. Higher confidence values indicate stronger predictions. Once SRA assigns returns risk scores, they are ranked from highest to lowest. W&I automatically selects and audits returns starting from the highest score and continues until it meets its predetermined audit workload.

IRS Has Not Comprehensively Reviewed Automated Systems to Mitigate Potential Biases

IRS officials stated that each year, managers, DDB programmers, and other IRS stakeholders meet to discuss DDB updates. During this meeting, they review the prior year's DDB performance, including the no-change rate, along with audit selection criteria. If they identify issues, they implement programming amendments. Furthermore, IRS stakeholders also review new legislation that could affect the refundable credits programs and lead to changes in its audit selection processes. IRS also documents these reviews with detailed meeting notes.

However, this review process does not consider all data inputs, assumptions, and other model components that could inform IRS about the demographic equity of the audit selection process. For example, while meeting notes indicate reviews of some rules, there is no evidence of a systematic, comprehensive review of the rules and filters contained in the DDB system.

We identified some components of W&I's automated audit selection process that could potentially skew selection toward returns with certain demographic characteristics that may not necessarily represent returns with the highest risk of noncompliance. For example, some risk scores contained in the DDB program vary by sex and have not been updated since the implementation of the DDB system in 2001.

In addition, the SRA model's calculation of the final risk score may lead to unintended results. For example, the SRA model contains 44 rules that predict a "change" to a tax return. Each rule is associated with a confidence value, a measure of the likelihood that the rule break accurately predicts noncompliance. The final risk score is based on the total of the confidence values of triggered rules and the total number of rules triggered. This method means that under certain conditions, one tax return could have a lower risk score than the another while triggering all of the same rules and more. According to IRS officials, the SRA model was implemented in 2009, and IRS replaced it with a new model in 2024.

As noted earlier, current research efforts within IRS have also estimated the existence of racial disparities in audits and have identified algorithmic biases as a potential source of the disparities.[36] Specifically, that research noted outdated models as possible contributions to algorithmic bias and, consequently, racial disparities in audits. IRS said that it plans to expand the use of AI and other advanced analytic methods to create new and improved case selection models.[37]

Multiple AI accountability frameworks explain the importance of continuous monitoring to ensure AI is reliable, relevant, and does not create or further advance bias.[38] Our AI accountability framework explains

[36]Hertz, T. et al., "Differences in Audit Rates by Race."

[37]Initiative 3.1 in the *IRS Inflation Reduction Act Strategic Operating Plan* commits the agency to using improved analytics to aid in the selection of cases predicted to be at risk of noncompliance.

[38]GAO-21-519SP. Principles developed by industry organizations include, among others, the "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms" from Fairness, Accountability, and Transparency in Machine Learning. Government agencies have also developed AI fairness principles or initiatives, such as Equal Employment Opportunity Commission's Artificial Intelligence and Algorithmic Fairness Initiative.

that a variety of monitoring activities should be employed to ensure AI systems function as intended. For example, monitoring should include reviews of statistical properties of the input data used to develop models. These reviews should also compare the relationship between the data inputs and the prediction outputs to determine the accuracy and utility of AI systems. In addition, internal control principles direct agency management to design control activities to achieve objectives and respond to risks. These control activities could be achieved through internal guidance used to ensure IRS fulfills its mission, strategic plan, goals, and objectives.

Although IRS conducts and documents a review of its automated audit selection process, IRS does not systematically check that the reviews are adequately considering model inputs. Moreover, IRS has not developed clear guidance for reviewing the DDB system for potential demographic biases. As IRS expands its use of AI and other advanced analytic models, employing robust systematic reviews could help ensure the technology is accurate and credible and does not result in unintentional bias by demographic characteristics. For example, IRS could include methods in its reviews to identify components of the system that have not been updated within a specified time frame. Meanwhile, clear guidance for reviewing the DDB system for potential demographic biases could help IRS assess whether its audit selection system is equitable for different demographic groups.

IRS Does Not Use Some Available Performance Measures to Review the Effectiveness of Automated Systems

We found opportunities for IRS to improve its use of performance measures in assessing the effectiveness of the DDB program. IRS reviews the DDB rules and filters annually and maintains detailed documentation of the topics it reviews. However, the primary measure for assessing the performance of the DDB system is the no-change rate. That rate, as discussed earlier, could be lower with the inclusion of audits concluded based on taxpayer nonresponse, which IRS research has found is higher among Black taxpayers.

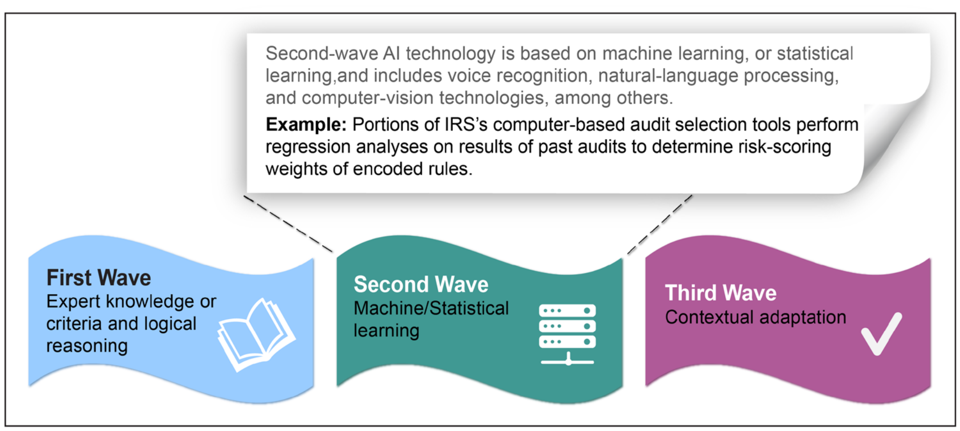
IRS provided documentation to support the validity of different DDB program rules. While that documentation included an assessment of the rules’ applicability and associated no-change rates, it was conducted in 2001 and did not provide any justifications for the original generation of the rules used. IRS officials said most of the rules were implemented to validate eligibility conditions of relationship, residency, and duplicate dependent claims. They also said those rules have not changed since 2001 and that the performance of those rules has been regularly evaluated over time.

IRS officials stated that the agency's Office of Research, Applied Analytics and Statistics and W&I's Return Integrity & Compliance Services began developing a new DDB scoring model in 2020. IRS intends for the new model to improve the detection of noncompliance in tax return filings. IRS officials said it was deployed in calendar year 2024, replacing the SRA model. Results from IRS's pilot of the new model indicate that it identifies largely different returns to audit compared to the existing model. As part of the pilot, IRS also considers the amount of revenue gained through conducting the audits as a performance measure, which is consistent with measuring the extent of noncompliance rather than just its presence.

While the results from the pilot indicate a lower no-change rate, the model also selected a high share of default audits, which include audits closed due to taxpayer nonresponse. As we noted earlier, default audits could lower the no-change rate. Further, IRS research indicated that taxpayer nonresponse is higher for Black taxpayers, indicating the new scoring model may still reflect biases.

The model will rely on modern machine learning processes, a marker of second wave AI systems (see fig. 4). According to IRS officials, modern machine learning methods offer the promise of increased accuracy, which will improve W&I's ability to distinguish compliant from noncompliant taxpayers. W&I officials said they plan to consider different factors in reviewing the new model. IRS officials also said they are researching best practices in algorithmic fairness literature.

Figure 4: The Second Wave of Artificial Intelligence (AI) in Internal Revenue Service (IRS) Audits



Source: GAO analysis of IRS information. | GAO-24-106126

In the case of broad agency goals, like those set forth in IRS's strategic operating plans, our past work on evidence-based policymaking has established that agencies should define goals that communicate the results that an organization seeks to achieve. These goals guide the organization's activities, and allow decision-makers, staff, and stakeholders to assess performance by comparing planned and actual results.[39] One key practice agencies should undertake is to define performance goals. These goals are generally expressed as tangible, measurable objectives, or as quantitative standards, values, or rates.

IRS relies primarily on the no-change rate to review the effectiveness of its automated audit selection process. It does not consider other available performance measures, which could increase the rigor of the review of its automated audit selection process. It also misses an opportunity to improve the likelihood that IRS is properly identifying returns at the highest risk of noncompliance and that the selection process does not unintentionally introduce demographic bias. More broadly, these improvements could allow IRS to assess with confidence that the system's components are valid.

IRS Does Not Have a Regular Process to Evaluate Reliability of External Databases Used in Audit Selection

IRS's automated system for audit selection may be using unreliable external data or data that have limited usefulness for assessing noncompliance risk. IRS cannot assure the data's reliability or usefulness because it has not conducted recent evaluations of the external databases housing these data. For instance, IRS does not contact state-level child support enforcement agencies to confirm information contained in the Federal Case Registry (FCR). However, if taxpayers choose to provide updated child support information to IRS, IRS will tell them to contact the state-level child support enforcement agencies that provided the information to the FCR. FCR databases may not represent current residency and child support details accurately, especially because taxpayers themselves must submit any changes.

The Kidlink database, another dataset used in IRS's automated audit selection processes, provides data on the child's name, the birth mother's Social Security number and name, and the birth father's Social Security number and name. Because it only includes parent-child relationships, it cannot directly verify all eligible relationships.[40] In addition, data in Kidlink were generally only available for children born after 1998 and could have

[39]GAO-23-105460.

[40]IRS officials said the Kidlink database is occasionally updated in the case of adoption.

affected the quality of selection of audits prior to 2015. IRS officials told us the agency is working with the Social Security Administration to fill in missing parental Social Security numbers. They said this effort should improve the accuracy of audit selection models.

In 2011, IRS provided an assessment of the FCR database in response to our prior work.[41] That assessment found that the FCR database was not sufficiently reliable to allow math error authority—an exception to standard audit practices—to be used in freezing refunds identified as potentially noncompliant.[42] The study also indicated that there were geographic inconsistencies in data reported to FCR. However, IRS continues to use the FCR database for pre-refund EITC audit selection. According to IRS officials, FCR still can provide data that can be used as a strong indicator of risk. However, they said FCR data are not substantive enough to justify the exercise of math error authority.

IRS's current research efforts on racial disparities in audits identified residency and relationship tests as potential sources of disparity. Specifically, the residency and relationship status of dependents must be imputed from incomplete information, and this process can introduce error. IRS researchers noted that imputation errors appear to raise the audit risk for Black EITC claimants relative to others. The researchers suggested this problem could be mitigated by modernizing audit selection models and supplementing existing data sources.

Standards for Internal Control in the Federal Government directs agency management to use quality information to achieve objectives. Data used by agencies should come from reliable sources and be free from error or bias. Management should evaluate data to ensure they meet these standards. If IRS were to assess the reliability of the Kidlink and FCR databases used in audit selection algorithms and document the results of its assessment, it could better understand how those databases affect the noncompliance risk assigned to some tax returns. Then IRS would have

[41]GAO, *Tax Administration: IRS's 2008 Filing Season Generally Successful Despite Challenges, although IRS Could Expand Enforcement during Returns Processing*, GAO-09-146 (Washington, D.C.: Dec. 12, 2008).

[42]As an exception to standard deficiency procedures, 26 U.S.C. § 6213(b)(1) authorizes IRS to summarily assess and collect tax without first providing the taxpayer with a notice of deficiency or access to the Tax Court when addressing "mathematical and clerical" errors (known as "math error authority").

information that would be useful for adjusting risk scores to reflect less reliable sources.

Conclusions

Recent research identified disparities in IRS audits by race. IRS analysts have confirmed these analyses. Audit disparities were largely estimated within EITC audits, indicating potential biases in IRS's process for selecting returns claiming the EITC and other refundable credits for audit. IRS has stated its commitment to administering the tax code fairly.

We found limitations in IRS's oversight of W&I's audit selection systems and, consequently, in IRS's ability to address equity—racial and otherwise—in audit selection. We noted difficulties in interpreting one of IRS's key goals for audit selection—a low no-change rate. IRS has neither measured the no-change rate without default audits, nor has it separated default audits based on taxpayer nonresponse.

IRS has already begun conducting audit equity research by analyzing the demographic distribution of audits, but it has not incorporated this analysis into audit workplan development. Furthermore, IRS does not have clear guidance for considering demographic equity in developing the audit workplan. This lack of guidance limits IRS's ability to determine the potential for biases that may be unintentionally embedded in its audit workplans and processes.

While IRS has a process for evaluating its primary system used to select refundable credits for audit—the DDB—it does not have a method for ensuring the components of the model are comprehensively reviewed. Further, IRS does not have guidance for assessing the automated audit selection system for its potential to contain algorithmic biases by demographic characteristics. IRS has already begun taking steps to address demographic equity in its audit selection processes, including plans to introduce more advanced artificial learning techniques into its systems. However, IRS does not have a comprehensive and systematic review process to help assure the accuracy, credibility, and demographic equity of those systems.

In addition, IRS's reviews of the DDB system's effectiveness rely primarily on one performance measure—the no-change rate. Consequently, the agency is limited in its ability to predict the risk of returns' noncompliance and to reduce potential biases in audit selection.

Finally, the DDB system relies on outside databases to identify potential noncompliance. IRS has not regularly and consistently assessed the reliability of those databases.

Recommendations for Executive Action

We are making six recommendations to IRS.

The Commissioner of Internal Revenue should direct W&I officials to calculate multiple no-change rates that include (1) the overall no-change rate, (2) the no-change rate excluding default audits, and (3) the no-change rate excluding taxpayer nonresponse default audits to provide insights into potential equity disparities. (Recommendation 1)

The Commissioner of Internal Revenue should direct W&I officials to develop guidance for considering audit equity research in developing the audit workplan. (Recommendation 2)

The Commissioner of Internal Revenue should direct W&I officials to establish a systematic process for ensuring its reviews of audit selection algorithms comprehensively consider data inputs, assumptions, and other model components. (Recommendation 3)

The Commissioner of Internal Revenue should direct W&I officials, as part of W&I's systematic review of its automated audit selection process, to develop clear guidance for assessing the DDB system's potential to contain algorithmic biases by demographic characteristics. (Recommendation 4)

The Commissioner of Internal Revenue should direct W&I officials to use additional performance measures, such as revenue protected, to assess the effectiveness of its automated audit selection system for predicting risk of noncompliance. (Recommendation 5)

The Commissioner of Internal Revenue should direct W&I officials to assess the reliability of the Kidlink and FCR databases used in audit selection algorithms and document the results of its assessment. If appropriate, W&I should then incorporate the results of its assessment into its audit selection algorithms. (Recommendation 6)

Agency Comments

We provided a draft of this report to IRS for review and comment. In its comments, reproduced in appendix II, IRS agreed with all of our recommendations, stating that it has created a team made up of subject matter experts from across IRS to address unintentional disparity. As part of this effort, IRS said it has created the Trustworthy Analytics team and the Advanced Analytics and AI Governance program to facilitate research and evaluation of fairness-related questions. IRS stated that it has already taken steps to improve fairness and equity in audit case selection, such as analyzing the EITC audit selection process to identify root causes of disparities.

In its response, IRS stated that it has established a process to review audit selection algorithms on a regular basis. We plan to review this process to ensure it satisfies our recommendation. IRS also noted that it currently uses revenue as another performance measure for evaluating its automated audit selection processes, and we are following up with the agency to determine whether its actions are sufficient to implement our recommendation. In addition, IRS described plans to address the other recommendations.

IRS also provided technical comments, which we incorporated as appropriate.

As agreed with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution until 26 days from the report date. At that time, we will send copies of this report to the appropriate congressional committees, the Commissioner of Internal Revenue, and other interested parties. In addition, the report will be available at no charge on the GAO website at https://www.gao.gov.

If you or your staff have any questions about this report, please contact me at (202) 512-6806 or lucasjudyj@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. GAO staff who made key contributions to this report are listed in appendix III.

Sincerely,

Jessica Lucas-Judy
Director, Tax Policy and Administration
Strategic Issues

Appendix I: Objectives, Scope, and Methodology

This report: (1) describes Internal Revenue Service (IRS) policies to address equity in audits; (2) describes relevant IRS audit planning and selection processes; (3) assesses equity considerations in IRS processes to develop audit workplans; and (4) analyzes how IRS addresses equity in its automated audit selection processes.

Recent research indicates there may be disparities in IRS audits by taxpayer demographics, such as race, sex, family composition, filing status, location, and income. One study provided direct evidence on racial disparities in IRS audits.[1] We conducted a detailed review of that study to assess its reliability. The authors presented validation analyses that supported the study findings. While we determined that analysis was reliable for the purpose of this report, we identified some potential limitations. The study relies on the Bayesian Improved First Name Surname and Geocoding method to predict the race of taxpayers. Government and academic researchers, including ourselves, have used this method or similar methods, which uses name and geolocation, to statistically predict a person's race. The authors used data on self-reported race from voter registries in North Carolina before 1993 to validate the BIFSG method. Weighting was used to align these data to national data. Still, the voter registry data include more limited definitions for race for older people, which might not be exchangeable for the race responses for all filers in 2014. Another potential limitation is that mortgage records were used for first names, and this selection bias could be correlated with race, due to differences in home ownership between the Black and non-Black populations. To the extent the BIFSG methodology is sensitive to these, the results may not apply to the full population of taxpayers.

To describe current IRS audit selection processes, we reviewed documentation on program procedures from the Internal Revenue Manual (IRM) and IRS documents describing audit procedures. We also interviewed IRS officials who oversee or who work on W&I's audit selection process, including officials in IRS's Small Business and Self-Employed division and analysts in IRS's Office of Research, Applied Analytics and Statistics.

For context, we obtained statistical information for fiscal year 2016 to 2023 about selected Wage and Investment Division (W&I) audit data.

[1]Elzayn, H. et al., *Measuring and Mitigating Racial Disparities in Tax Audits,* Stanford Institute for Economic Policy Research (Jan. 30, 2023).

That information included data on pre- and post-refund audits, audits by type of return, audits by selection method, and default audits, which are audits closed due to insufficient taxpayer response. Those data came from IRS's Dependent Database (DDB) tables and IRS workplans, among other sources.

To analyze how W&I's procedures for selecting tax returns for audit address equity, we reviewed W&I's audit selection procedures and related internal controls intended to help W&I achieve its stated mission of "applying the tax law with integrity and fairness to all."[2] We then assessed whether these W&I procedures followed selected internal control standards from *Standards for Internal Control in the Federal Government*.[3] We also assessed these procedures against relevant agency goals and our past work on key practices in evidence-based policy making and performance management. Furthermore, we interviewed relevant IRS officials concerning their understanding of W&I's mission, objectives, and internal controls related to W&I's audit selection.

In addition, we reviewed the Department of the Treasury's Strategic Plan Fiscal Year 2022-2026, IRS's Strategic Plan FY 2022-2026, IRM sections related to W&I's mission statement and program objectives, IRS's public statements for audits, and other IRS documentation related to W&I's audit selection process, including its Enterprise Scenario Planning Tool. We also conducted a detailed review of IRS's DDB Functional Specification Package. As part of this review, we consulted with internal experts on statistical learning programs and data analysis, and our Accountability Framework for Artificial Intelligence.[4]

We also determined that the internal controls principles related to control activities and information and communication were significant to the objective looking at automated selection processes. These principles were that management should design control activities to achieve objectives and respond to risks and use quality information. We assessed whether W&I's audit selection process addressed the selected internal control standards by reviewing documentation and interviewing IRS officials familiar with the selection process. The controls we chose applied

[2]IRM § 1.1.13.1.1(1) (10-13-2021).

[3]GAO-14-704G.

[4]GAO, *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities*, GAO-21-519SP (Washington, D.C.: June 2021).

to the processes and procedures that W&I officials described or for which they provided documentation regarding how audits were selected.

We conducted this performance audit from June 2022 to April 2024 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Appendix II: Comments from the Internal Revenue Service

DEPARTMENT OF THE TREASURY
INTERNAL REVENUE SERVICE
WASHINGTON, DC 20224

DEPUTY COMMISSIONER

April 11, 2024

Ms. Jessica Lucas-Judy
Director, Strategic Issues
U.S. Government Accountability Office
441 G Street, N.W.
Washington, DC 20548

Dear Ms. Lucas-Judy:

Thank you for the opportunity to review and comment on the draft report entitled Tax Enforcement: IRS Audit Selection Processes for Returns Claiming Refundable Credits Could Better Address Equity (GAO-24-106126).

The report points to research that indicates disparities by taxpayer demographics, such as race, sex, family composition, filing status, location, and income, may exist in Internal Revenue Service (IRS) audit processes. The IRS is committed to enforcing the tax laws in a manner that is fair and impartial. It is important to reiterate that the IRS does not and will not consider race in our case selection and audit processes. We are committed to improving fairness in tax administration. This commitment is reflected in the IRS' Inflation Reduction Act Strategic Operating Plan (IRA SOP) for FY 2023 to 2031.

As Commissioner Werfel outlined in his letter to the Senate Committee on Finance, dated September 18, 2023, the IRS has already taken significant steps to improve fairness in tax administration by realigning resources internally to accelerate progress against our strategic priorities. Starting in fiscal year 2024, we have substantially reduced the number of correspondence audits focused specifically on certain refundable tax credits, including Earned Income Tax Credit, American Opportunity Tax Credit, Health Insurance Premium Tax Credit and Additional Child Tax Credit. Relying too heavily on audits to resolve basic errors can lead to fewer taxpayers receiving credits and deductions to which they are entitled and thus decrease accuracy in tax administration. Focusing on helping taxpayers submit accurate filings upfront will increase payment accuracy while reducing administrative burdens for the IRS and the tax filers. We are also increasing our efforts to address unscrupulous preparers who are leading their customers to underreport income or overclaim credits and deductions.

2

We also have created a team made up of subject matter experts from across the IRS to address unintentional disparity. The Exam Disparity team has already taken steps to improve fairness and equity in audit case selection. In addition to reducing the number of EITC audits, the team convened to further explore the unintentional disparities. In addition, we created the Trustworthy Analytics team and the Advanced Analytics and AI Governance program to facilitate research and evaluation of fairness-related questions. Accomplishments to date include:

- We implemented a new scoring model for EITC audit section, which was initially developed and tested in 2020-2022. Prior to implementation, we documented that it is expected to reduce racial disparity among the taxpayers it selects.
- We initiated two new pilot programs whose goals are to increase audit revenue, reduce burden on compliant taxpayers, and reduce racial disparity by developing new audit selection models that rely on state-of-the-art machine learning technology.
- Leveraging best practices from the field of data science, we are developing a governance process to evaluate all enforcement models across the IRS, in terms of fairness, transparency, replicability, and capability of being explained.
- We conducted an in-depth analysis of the EITC audit selection process that identified many of the root causes of disparity in audit selection and are taking action to address those.
- We continued to improve our methods for estimating outcomes for different demographic groups by using established techniques that do not rely upon taxpayers' actual race/ethnicity information.

Through our parallel efforts to help taxpayers “get it right” through improved service, outreach and education, and other tools, coupled with our increase focus on addressing unscrupulous return preparers and other bad actors, we are taking new steps to increase payment accuracy. We anticipate it will take several months after the end of the filing season to determine the effect of these changes, but we are committed to monitoring, sharing our findings and making additional improvements.

We agree with all of GAO’s findings, which will assist us in continuing to assess and address disparities. As we continue to review and revise our case selection process, we will work to identify any disparities across dimensions including age, gender, geography, race and ethnicity as well as continual refining our approaches to compliance and enforcement to improve fairness in tax administration and maintain accountability to taxpayers as informed by our research.

We appreciate GAO recognizing that we have already taken steps to address demographic equity in our audit selection processes, including introducing advanced artificial learning techniques into our systems. We will continue our efforts to address equity in correspondence audits.

3

The IRS remains committed to enforcing our nation's tax laws with integrity and fairness to all. Our responses to your specific recommendations are enclosed. If you have any questions, please contact James Clifford, Director, Return Integrity and Compliance Services, Wage and Investment Division, at (470) 639-3250.

Sincerely,

Douglas W. Odonnell Digitally signed by Douglas W. Odonnell Date: 2024.04.11 15:18:22 -04'00'

Douglas W. O'Donnell
Deputy Commissioner

Enclosure

Enclosure

Recommendations for Executive Action

We are making six recommendations to IRS.

RECOMMENDATION 1
The Commissioner of Internal Revenue should direct W&I officials to calculate multiple no-change rates that include (1) the overall no-change rate, (2) the no-change rate excluding default audits, and (3) the no-change rate excluding taxpayer nonresponse default audits to provide insights into potential equity disparities.

RESPONSE
We agree with this recommendation. We will include additional calculations of the no change rate in the analysis of the performance of the rules and filters at the annual Dependent Database (DDB) meetings starting in 2024.

RECOMMENDATION 2
The Commissioner of Internal Revenue should direct W&I officials to develop guidance for considering audit equity research in developing the audit workplan.

RESPONSE
We agree with this recommendation. While the IRS does not and will not use race data in our audit selection processes, the IRS Strategic Operating Plan commits the IRS to developing service-wide policies regarding the appropriate and legal means available to mitigate audit rate disparities.

RECOMMENDATION 3
The Commissioner of Internal Revenue should direct W&I officials to establish a systematic process for ensuring its reviews of audit selection algorithms, comprehensively consider data inputs, assumptions, and other model components.

RESPONSE
We agree with the recommendation. We have established a process to review audit selection algorithms on a regular basis.

RECOMMENDATION 4
The Commissioner of Internal Revenue should direct W&I officials, as part of W&I's systematic review of its automated audit selection process, to develop clear guidance for assessing the DDB system's potential to contain algorithmic biases by demographic characteristics.

RESPONSE
We agree with the recommendation. This commitment is articulated in the IRS-wide policies and procedures that are in development, as referenced in our responses to

2

Recommendations 2 and 3. An analysis will be completed post-selection to identify potential impact of demographic inequity in the audit selection.

RECOMMENDATION 5

The Commissioner of Internal Revenue should direct W&I officials to use additional performance measures, such as revenue protected, to assess the effectiveness of its automated audit selection system for predicting risk of noncompliance.

RESPONSE

We agree with the recommendation. We currently evaluate the effectiveness of our audit selection based on both no change rates and revenue for predicting the risk of noncompliance and will continue to do so. During the annual Dependent Database (DDB) meeting, we review both the no change rate and revenue protected for the various filters and rules. They are key metrics used to evaluate the performance of models and are reflected in the program's quarterly reports along with other performance metrics.

RECOMMENDATION 6

The Commissioner of Internal Revenue should direct W&I officials to assess the reliability of the Kidlink and FCR databases used in audit selection algorithms and document the results of its assessment. If appropriate, W&I should then incorporate the results of its assessment into its audit selection algorithms.

RESPONSE

We agree with the recommendation. We will explore assessing the reliability of the Kidlink database used in audit selections. The IRS regularly updates Kidlink to enhance the reliability of those data. There is already a process for documenting the reliability of the Federal Case Registry (FCR) database used in audit selections. We will explore further assessment of the FCR database.

Appendix III: GAO Contact and Staff Acknowledgments

GAO Contact

Jessica Lucas-Judy, (202) 512-6806 or lucasjudyj@gao.gov

Staff Acknowledgments

In addition to the contact named above, Sonya Phillips (Assistant Director), Jennifer G. Stratton (Analyst-in-Charge), Jieun Chang, Caitlin Cusati, Robert Gebhart, Krista Loose, Edward Nannenhorn, Robert Robinson, Shep Ryen, Jeff Tessin, Sonya Vartivarian, Peter Verchinski, and John Villecco made key contributions to this report.

GAO's Mission

The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through our website. Each weekday afternoon, GAO posts on its website newly released reports, testimony, and correspondence. You can also subscribe to GAO's email updates to receive notification of newly posted products.

Order by Phone

The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, https://www.gao.gov/ordering.htm.

Place orders by calling (202) 512-6000, toll free (866) 801-7077, or
TDD (202) 512-2537.

Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.

Connect with GAO

Connect with GAO on Facebook, Flickr, Twitter, and YouTube.
Subscribe to our RSS Feeds or Email Updates. Listen to our Podcasts.
Visit GAO on the web at https://www.gao.gov.

To Report Fraud, Waste, and Abuse in Federal Programs

Contact FraudNet:

Website: https://www.gao.gov/about/what-gao-does/fraudnet

Automated answering system: (800) 424-5454 or (202) 512-7700

Congressional Relations

A. Nicole Clowers, Managing Director, ClowersA@gao.gov, (202) 512-4400, U.S. Government Accountability Office, 441 G Street NW, Room 7125, Washington, DC 20548

Public Affairs

Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, DC 20548

Strategic Planning and External Liaison

Stephen J. Sanford, Managing Director, spel@gao.gov, (202) 512-4707
U.S. Government Accountability Office, 441 G Street NW, Room 7814,
Washington, DC 20548