United States Government Accountability Office

Report to the Ranking Member, Committee on Ways and Means, House of Representatives

May 2024

TAX GAP

IRS Should Take Steps to Ensure Continued Improvement in Estimates

GAO-24-106449

GAO Highlights

Highlights of GAO-24-106449, a report to the Ranking Member, Committee on Ways and Means, House of Representatives

May 2024

TAX GAP

IRS Should Take Steps to Ensure Continued Improvement in Estimates

Why GAO Did This Study

The tax gap—the difference between the tax amounts voluntarily paid by individuals and businesses and the tax amounts that are owed—has been a persistent problem for decades.

The tax gap is an aggregate estimate of individual income, corporation income, employment, and estate taxes. IRS estimates are based on three types of noncompliance: (1) underreporting of tax liabilities on timely filed tax returns; (2) underpayment of taxes due from timely filed returns; and (3) nonfiling, when a taxpayer fails to file a required tax return altogether or on time.

GAO was asked to review IRS's most recent tax gap estimates for tax years 2014-2016. This report (1) describes IRS's estimates and changes from prior estimates; (2) assesses IRS's changes to the sample design of NRP; and (3) assesses how the estimates help IRS and other stakeholders gain insights into noncompliance. GAO reviewed IRS's tax gap data and reports and interviewed IRS officials.

What GAO Recommends

GAO is making six recommendations to IRS, including that IRS conduct additional analyses to understand the root causes of undetected noncompliance and complete its documentation related to the pilot process for sampling returns.

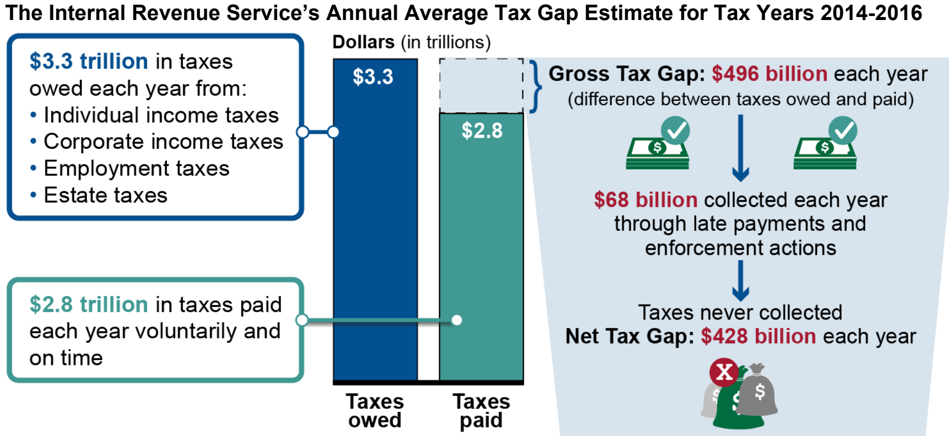
IRS agreed with all six of GAO's recommendations and described steps it plans to take in response to each recommendation.

View GAO-24-106449. For more information, contact James R. McTigue at (202) 512-6806 or McTigueJ@gao.gov.

What GAO Found

Taxpayers voluntarily and timely paid about 85 percent of the taxes they should have paid for tax years 2014-2016, according to the Internal Revenue Service's (IRS) latest tax gap estimate. IRS most recently projected the tax gap will grow to $688 billion for tax year 2021. However, when measured relative to the overall economy, the tax gap has remained relatively stable.

The Internal Revenue Service's Annual Average Tax Gap Estimate for Tax Years 2014-2016



Source: Internal Revenue Service (IRS). | GAO-24-106449

In developing the estimates, IRS applies a statistical technique to National Research Program (NRP) audit data to account for noncompliance that was not detected by examiners in audits. Applying this technique nearly doubles the individual underreporting tax gap estimate, increasing the estimate from $145 billion to $278 billion. IRS has not conducted analysis to understand the causes of this estimate of undetected noncompliance. By doing so, IRS would be better positioned to help improve the reliability of and confidence in the adjustment and potentially examiners' detection of noncompliance.

IRS is piloting a new process for sampling tax returns for NRP's audits. The new process uses artificial intelligence (AI) to improve the efficiency and selection of audit cases to help identify noncompliance. However, IRS has not completed its documentation of several elements of its AI sample selection models, such as key components and technical specifications. Completing documentation would help IRS retain organizational knowledge, ensure the models are implemented consistently, and make the process more transparent to future users.

IRS's Strategic Operating Plan (SOP) describes how IRS intends to spend the tens of billions of dollars it received under the Inflation Reduction Act. However, the SOP is not clearly linked to tax gap data. For example, the plan does not address sole proprietor noncompliance, which is one of the largest areas of tax noncompliance. Furthermore, IRS did not address a prior GAO recommendation to link NRP data to its compliance efforts. As IRS develops implementation plans for the SOP initiatives and projects, it has an opportunity to further integrate a significant source of evidence to refine its compliance strategies, thereby addressing the recommendation. Linking compliance strategies with data would help provide assurance that IRS is effectively allocating enforcement resources.

Contents

Letter		1
	Background	4
	IRS's Tax Gap Estimate Remains Stable, but IRS Has Not Examined Root Causes of Undetected Noncompliance	10
	IRS Is Using Artificial Intelligence Models for NRP Audit Selection, but Does Not Have Controls Fully Aligned with Key Accountability Practices	23
	IRS Seeks to Identify Emerging Issues, but Could Do More to Link Tax Gap Data to Its Strategic Operating Plan	32
	Conclusions	40
	Recommendations for Executive Action	41
	Agency Comments and Our Evaluation	42
Appendix I	Objectives, Scope, and Methodology	44
Appendix II	Additional Information on Undetected Noncompliance in Tax Gap Estimates	48
Appendix III	National Research Program Redesign Operational Changes	50
Appendix IV	Comments from the Internal Revenue Service	51
Appendix V	GAO Contact and Staff Acknowledgments	56
Tables		
	Table 1: Estimated Average Annual Individual Income Underreporting Tax Gap by Tax Return Item or Category, Tax Years 2014-2016	14
	Table 2: National Research Program (NRP) Case Counts and Changes to NRP by Tax Year, 2001-2021	23
	Table 3: Internal Revenue Service (IRS) Could Better Document Its Efforts to Implement National Research Program (NRP) Artificial Intelligence (AI) Models	28

Table 4: External Stakeholders' Suggestions for Improving Tax Gap Estimates 35
Table 5: Estimated Average Annual Individual Income Underreporting Tax Gap by Tax Return Item or Category, Tax Years 2014-2016 48
Table 6: Internal Revenue Service (IRS) Efforts to Achieve the Tax Year 2019 National Research Program (NRP) Redesign Goals 50

Figures

Figure 1: IRS's Annual Average Tax Gap Estimate for Tax Years 2014-2016 4
Figure 2: Internal Revenue Service Estimates and Projections of Gross Tax Gap 5
Figure 3: Given Available Data, the IRS Cannot Fully Represent Noncompliance in Some Components of the Tax System 6
Figure 4: Net Misreporting as Percentage of Total Income, by Third-Party Information Reporting, Tax Years 2014-2016 9
Figure 5: IRS Estimates of Gross Tax Gap and Taxes Paid Voluntarily and Timely in Dollars and as a Percent of Gross Domestic Product (Tax Years 2001-2021) 11
Figure 6: Estimated Average Annual Gross Tax Gap by Type of Noncompliance and Tax, Tax Years 2014-2016 13
Figure 7: Estimated Misreporting of Sole Proprietor Income, Annual Average Tax Years 2014-2015 16
Figure 8: Income before and after IRS Audit for Taxpayers with Losses or No Income, Annual Average Tax Years 2014-2015 18
Figure 9: Traditional and Redesigned National Research Program (NRP) Sample Selection Process 26

Abbreviations

AI	artificial intelligence
AI Framework	AI Accountability Framework
DCE	detection controlled estimation
EITC	Earned Income Tax Credit
IRA	Inflation Reduction Act
IRS	Internal Revenue Service
GDP	gross domestic product
JCT	Joint Committee on Taxation
NRP	National Research Program
RAAS	Research, Applied Analytics and Statistics
SB/SE	Small Business/Self Employed
SOP	Strategic Operating Plan
TIGTA	Treasury Inspector General for Tax Administration

U.S. GOVERNMENT ACCOUNTABILITY OFFICE

441 G St. N.W.
Washington, DC 20548

May 6, 2024

The Honorable Richard E. Neal
Ranking Member
Committee on Ways and Means
House of Representatives

Dear Mr. Neal,

Every year, taxpayers fail to pay hundreds of billions of dollars in taxes. The tax gap—the difference between taxpayers' true tax liability and what they paid on time—has been a persistent problem for decades. In October 2022, the Internal Revenue Service (IRS) released its most recent estimate of the tax gap, which covered tax years 2014-2016.[1] IRS estimated the average annual gross tax gap to be $496 billion for each of those years. IRS also estimated that $68 billion would be collected through enforcement actions or late payments, leaving a net annual tax gap of $428 billion.

In October 2023, IRS released projections of the tax gap for tax years 2020 and 2021. These projections assume that compliance rates found in the 2014-2016 audits continue in 2020 and 2021. For tax year 2021, IRS projected a gross tax gap of $688 billion and net tax gap of $625 billion.

IRS develops the tax gap estimate using data from several sources including its National Research Program (NRP), which conducts audits of a random sample of tax returns. According to IRS, tax gap estimates provide it with periodic insights into the nature and extent of noncompliance for use in formulating tax administration strategies. Meanwhile, other stakeholders, such as congressional committees, use the tax gap estimate to inform discussions of proposed policies and changes to the law, among other purposes.

[1]Internal Revenue Service, *Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2014-2016*, Publication 1415 (Rev. 10-2022) (Washington, D.C.: October 2022).

In 2022, Congress provided IRS tens of billions of dollars through the Inflation Reduction Act (IRA), in part, to increase enforcement.[2] According to IRS's IRA Strategic Operating Plan (SOP), some of the funds will be used for initiatives aimed at reducing the tax gap, researching areas of noncompliance that the tax gap estimate previously has not covered, and improving the overall estimate.[3] IRS also plans to use the funds to improve services that help taxpayers meet their obligations, among other things. According to the SOP, IRS will know it has been successful if the tax gap decreases relative to the size it would have been without IRA funding.

You asked us to review IRS's latest tax gap estimate for tax years 2014-2016. This report (1) describes the main types of noncompliance in IRS's estimate of the gross and net tax gap for tax years 2014-2016 and changes from prior estimates, (2) assesses IRS's planned changes to the NRP, including any changes to sample design and use of artificial intelligence; and (3) assesses the extent to which the estimate helps IRS and other stakeholders gain insights into noncompliance for formulating tax administration strategies.

To address these objectives, we reviewed IRS's tax gap estimate and projections and IRS technical documents on estimating compliance and the tax gap and ensuring taxpayer compliance. We also reviewed our past reports and Treasury Inspector General for Tax Administration reports on IRS's tax gap estimate. We examined data from IRS's NRP study of individual tax returns for tax years 2014 and 2015 to provide additional detail on aspects of taxpayer compliance. We also interviewed IRS officials responsible for estimating the tax gap.

We reviewed IRS documentation and academic papers on the tax gap estimate, projections, and NRP data and determined that they were sufficiently reliable for the purposes of reporting on the components of the tax gap and trends over time. We discuss a key limitation in the report related to the statistical adjustment that accounts for half of the size of the

[2]Pub. L. No. 117-169, tit. I, subtit. A, pt. 3, § 10301(1), 136 Stat. 1818, 1831–1832 (2022). In June 2023, the Fiscal Responsibility Act of 2023 rescinded about $1.4 billion of amounts appropriated for IRS by IRA. See Pub. L. No. 118-5, div, B, tit. II, § 251, 137 Stat. 10, 30-31 (2023). In March 2024, the Further Consolidated Appropriations Act of 2024 rescinded another $20.2 billion of amounts appropriated for IRS by IRA. See Pub. L. No. 118-47, div, B, tit. VI, § 640, div, D, tit. V, § 530, H. R. 2882, at 113, 249 (2024).

[3]Internal Revenue Service, *Internal Revenue Service Inflation Reduction Act Strategic Operating Plan*, Publication 3744 (Rev. 4-2023) (Washington, D.C.: April 2023).

tax gap. We compared the methodology used by IRS to estimate undetected noncompliance to *Standards for Internal Control in the Federal Government* on quality information use and Office of Management and Budget information quality guidelines.[4]

We reviewed IRS documentation on its NRP redesign pilot, including technical specifications related to changes to the sample design, the use of artificial intelligence, and evaluation plans related to the sample selection pilot. We interviewed IRS officials responsible for the NRP redesign efforts. We compared IRS efforts to *Standards for Internal Control in the Federal Government* for documentation, selected practices in our Artificial Intelligence Framework, and leading practices for pilot design.[5]

We interviewed a nongeneralizable selection of external stakeholders who use the tax gap estimate to determine how useful the estimate is for their purposes and areas for potential improvement. The stakeholders included staff from seven congressional committees who have an interest in tax writing, as well as IRS funding or federal budget oversight. We interviewed other external stakeholders from the National Taxpayer Advocate Service, Internal Revenue Service Advisory Committee, and the Electronic Tax Administration Advisory Committee. We also interviewed IRS officials in its operating divisions and the Deputy Commissioner for Service and Enforcement. We reviewed IRS's Inflation Reduction Act Strategic Operating Plan to determine how IRS linked its plan to tax gap data.

See appendix I for more information on our objectives, scope, and methodology.

We conducted this performance audit from December 2022 to May 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

[4]GAO, *Standards for Internal Control in the Federal Government*, GAO-14-704G (Washington, D.C.: Sept. 10, 2014); and *Office of Management and Budget, Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by Federal Agencies; Republication*, 67 Fed. Reg. 8452 (Feb. 22, 2002).

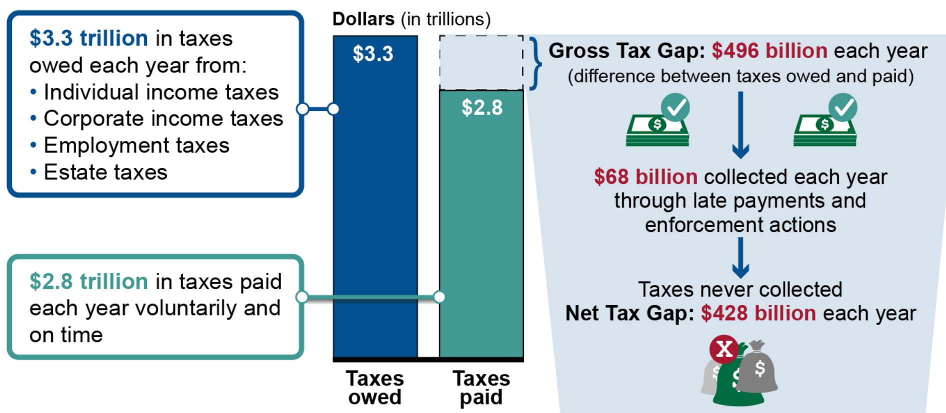
[5]GAO-14-704G; GAO, *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities*, GAO-21-519SP (Washington, D.C.: June 30, 2021); and *Data Act: Section 5 Pilot Design Issues Need to Be Addressed to Meet Goal of Reducing Recipient Reporting Burden*, GAO-16-438 (Washington, D.C.: Apr. 19, 2016).

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

IRS's most recent tax gap estimate found that taxpayers owed about $3.3 trillion per year in federal taxes for tax years 2014-2016. Taxpayers voluntarily paid approximately $2.8 trillion of that amount, resulting in a voluntary compliance rate of 85 percent. The remaining tax owed is the tax gap (see fig. 1).

Figure 1: IRS's Annual Average Tax Gap Estimate for Tax Years 2014-2016

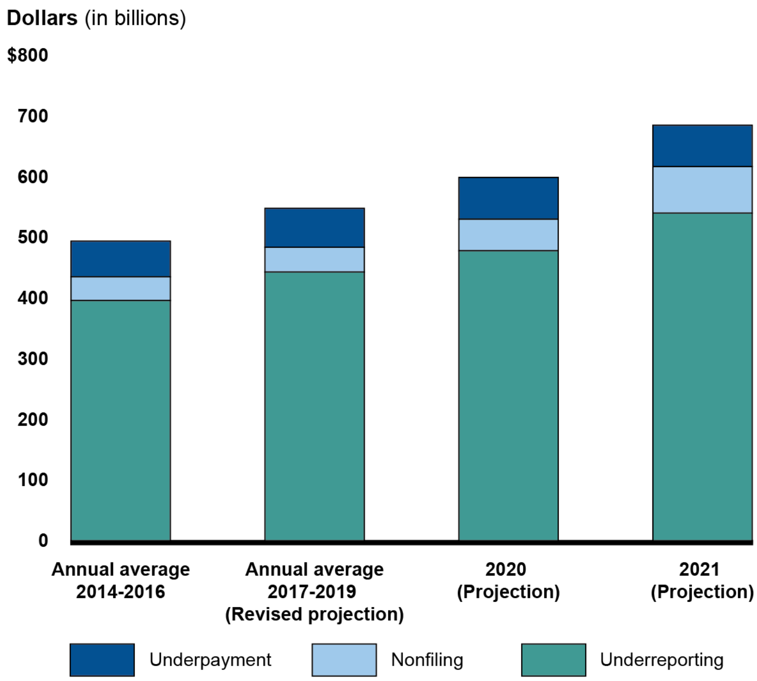


Source: Internal Revenue Service (IRS). GAO illustrations. | GAO-24-106449

The tax gap estimate is developed by IRS's Research, Applied Analytics and Statistics Division. The estimate for 2014-2016 was published in October 2022, which is 6 years after the last tax year covered by the estimate. The lag occurs because once individual taxpayers file their returns, IRS must select returns to be audited, complete the audits, and validate the compliance data collected before it can finalize the estimate.

To address interest for a more current estimate, IRS recently developed tax gap projections. According to IRS, these figures are called projections instead of estimates to highlight that projections are based upon taxpayer behavior observed in earlier years. IRS's most recent projections are for tax years 2020 and 2021. Generally, these projections assume compliance rates have not changed from the tax years 2014-2016 estimate. Applying these compliance rates to the tax year 2021 tax base, IRS projected the gross tax gap for that year was $688 billion (see fig. 2).

Figure 2: Internal Revenue Service Estimates and Projections of Gross Tax Gap



Source: GAO analysis of Internal Revenue Service Tax Gap estimates and projections. | GAO-24-106449

The gross tax gap estimate is an aggregate estimate of the different types of taxes that IRS administers—individual income, corporation income, employment, and estate.[6] For each tax type, IRS attempts to estimate the tax gap based on three types of noncompliance: (1) underreporting of tax liabilities on timely filed tax returns; (2) underpayment of taxes due from timely filed returns; and (3) nonfiling, when a taxpayer fails to file a required tax return altogether or on time.[7]

Noncompliance includes unintentional errors as well as intentional evasion by taxpayers, such as intentionally underreporting income, intentionally overreporting expenses, and engaging in abusive tax shelters or frivolous tax schemes. IRS has reported that the tax gap estimate can include a significant amount of noncompliance due to tax

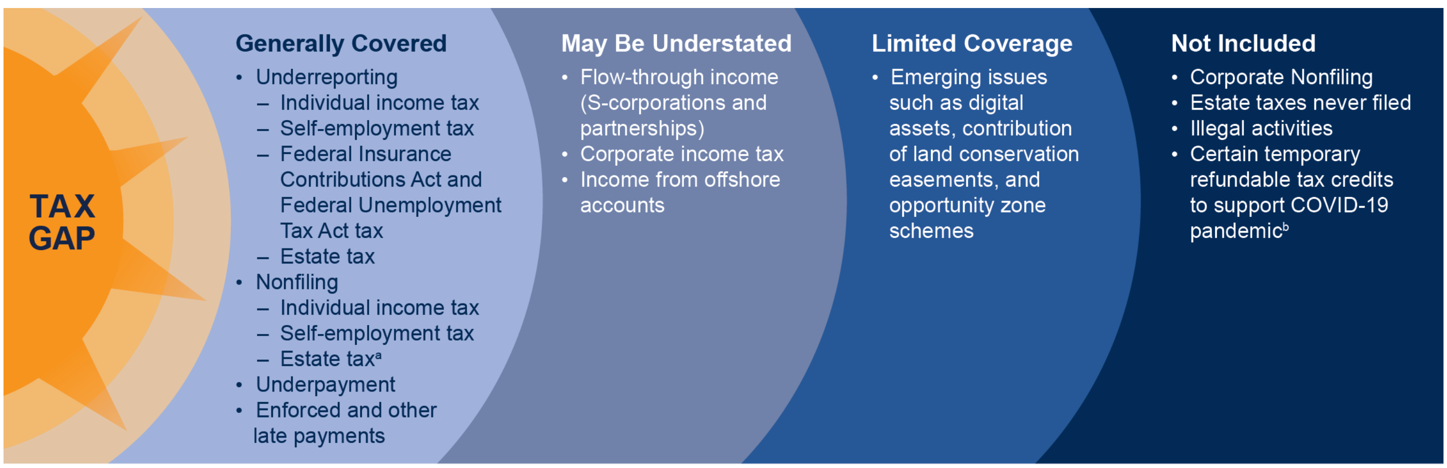
[6]Employment taxes include both employer-withheld employment taxes and self-employment taxes.

[7]By definition, legal tax avoidance (i.e., legally lowering tax liability) is not included in the tax gap.

law complexity that results in errors of ignorance, confusion, and carelessness. However, IRS does not have sufficient data to clearly distinguish the amount of noncompliance that arises from willful, as opposed to unintentional, mistakes. Moreover, the line between intentional and unintentional mistakes can be difficult to discern.

The tax gap estimate does not cover all areas of noncompliance. For example, IRS has been conducting research into estimating the corporate nonfiling tax gap but has not found a sufficiently reliable method to generate a robust estimate. IRS has also not developed an estimate for the tax gap related to income earned from illegal activities. IRS has reported that the tax gap associated with illegal activities is outside the scope of estimation because the overall goal of the government is to stop those activities, which would eliminate any associated tax (see fig. 3).

Figure 3: Given Available Data, the IRS Cannot Fully Represent Noncompliance in Some Components of the Tax System



Source: GAO analysis of Internal Revenue Service (IRS) information. | GAO-24-106449

[a]Estate nonfiling tax gap only reflects late filed returns. It does not include the tax gap associated with estate tax returns that are never filed.

[b]This is only applicable to tax year 2020 and 2021 projections.

In other areas, IRS lacks data to estimate noncompliance fully and reliably. For example, the tax gap estimate may be understated with respect to flow-through income (S-corporation and partnership income) and income from offshore accounts because of the difficulty in detecting sophisticated forms of noncompliance in audits.[8] IRS reports that the complex nature of these returns and the resources required to audit them suggests that even the top examiners might not be able to consistently detect all noncompliance. In emerging areas, such as noncompliance related to income from digital assets, IRS reported that it takes time to develop the expertise to uncover noncompliance and to complete examinations. In 2020, we recommended IRS take steps to increase third-party reporting on taxable transactions involving virtual currencies.[9] In response to our recommendation, Congress passed legislation requiring third party reporting on digital assets such as virtual currency.[10] Once implemented, this new information reporting should help IRS detect noncompliance in this emerging area.

IRS uses multiple approaches to estimate components of the tax gap. IRS reported that it generally used the same approaches to estimate the components of the tax years 2014-2016 estimate as it used in the tax years 2011-2013 estimate. IRS also reported it improved some methodologies. For example, IRS updated its methodologies for individual income tax and self-employment tax nonfiling to incorporate an expanded set of data.

IRS does not present standard errors, or any measures of uncertainty, associated with the estimates because the different methodologies used different standard errors and there are challenges with integrating standard errors across methods. However, IRS acknowledges that uncertainty exists.[11]

IRS's NRP audits a stratified random sample of individual income tax returns to provide a statistically valid representation of the compliance characteristics of these taxpayers. The individual income tax and self-employment tax gap estimates are based on these data. IRS

[8]Partnerships and S-corporations are flow-through entities—entities that generally do not pay taxes themselves on income, but instead, pass income or losses to their partners and shareholders, who must include that income or loss on their income tax returns. An S-corporation is a corporation meeting certain requirements that elects to be taxed under subchapter S of the Internal Revenue Code.

[9]GAO, *Virtual Currencies: Additional Information Reporting and Clarified Guidance Could Improve Tax Compliance*, GAO-20-188 (Washington, D.C.: Feb. 12, 2020).

[10]Pub. L. No. 117-58, div, H, tit. VI, § 80603, 135 Stat. 429, 1339–1341 (2021).

[11]IRS estimates of the tax gap may have different types of error, including sampling error, measurement error, estimation error, and coverage error. The projections of the tax gap estimates also have projection error.

subsequently applies a statistical technique to try to account for differences in examiner skills and experience. NRP data are also used to inform IRS decision-making and resource allocation.[12]

IRS has reported that the extent to which individual taxpayers accurately report their income is closely correlated with third-party reporting of income to taxpayers and IRS by third parties such as employers. IRS uses information reporting to verify data reported by taxpayers on their tax returns. According to IRS officials, information reporting also produces indirect benefits by increasing taxpayers' incentive to comply, because they know that IRS receives the information. We previously made recommendations to IRS to strengthen its information reporting programs.[13]

According to IRS's tax years 2014-2016 estimate, taxpayers misreport approximately 1 percent of income when substantial third-party information reporting exists. However, taxpayers misreport approximately 55 percent of income for which there is little or no third-party information reported to taxpayers and IRS, such as with sole proprietor income (see fig. 4).[14]

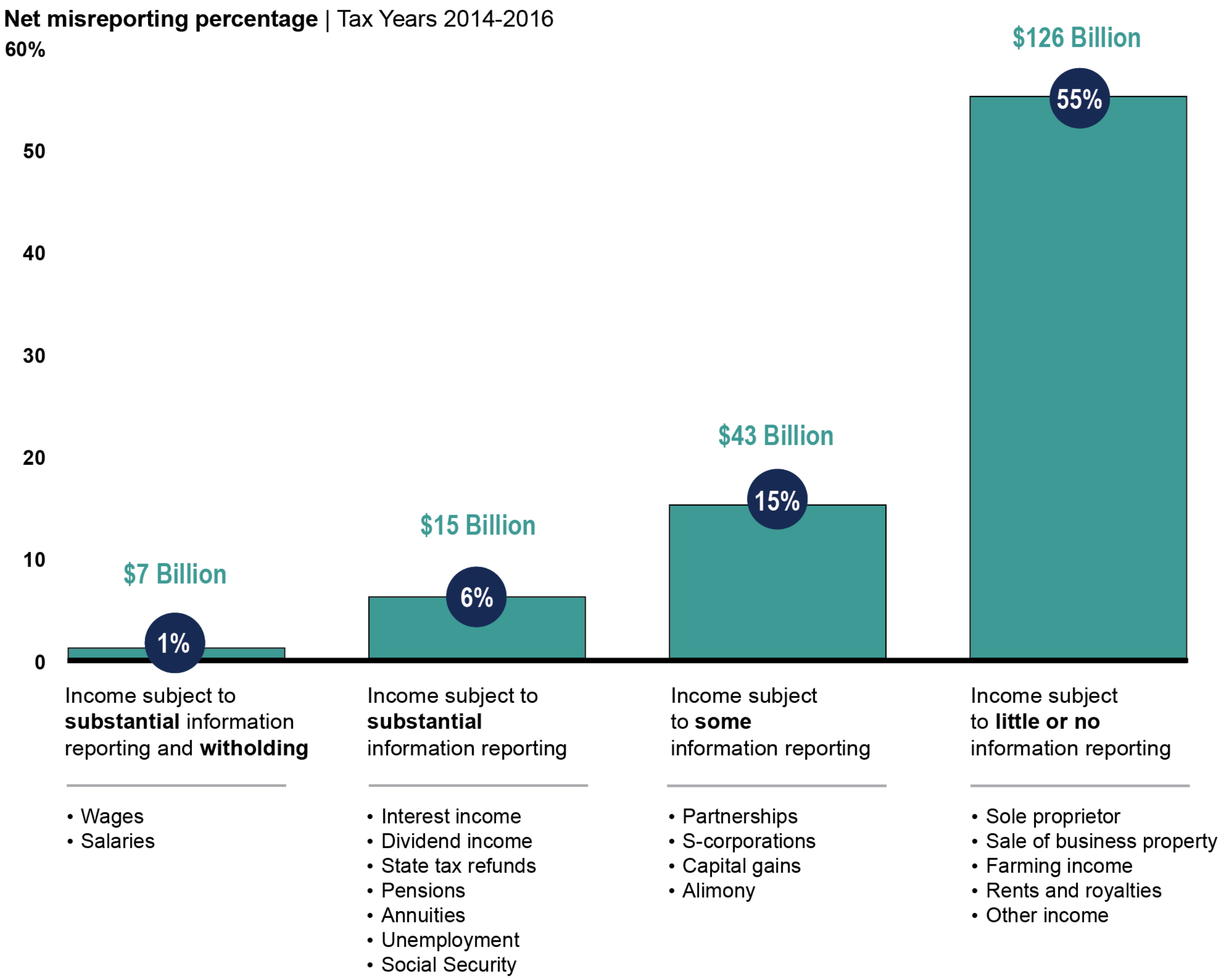
---

[12]Internal Revenue Manual, National Research Program Overview, § 4.22.1.1.

[13]GAO, *Tax Administration: Better Coordination Could Improve IRS's Use of Third-Party Information Reporting to Help Reduce the Tax Gap,* GAO-21-102 (Washington, D.C.: Dec. 15, 2020). As of January 2024, IRS implemented three of the nine recommendations. IRS neither agreed nor disagreed with the other six but has taken steps toward implementing them.

[14]A sole proprietor is an unincorporated business with one owner. Sole proprietors generally report their business-related profit or loss on their Schedule C, *Profit or Loss from Business*, which is filed as part of their individual income tax returns, IRS Form 1040. Examples of sole proprietors include skilled trades and professions such as plumbers, gig workers, and social media influencers, among many others.

Figure 4: Net Misreporting as Percentage of Total Income, by Third-Party Information Reporting, Tax Years 2014-2016



Source: GAO analysis of Internal Revenue Service (IRS) data. | GAO-24-106449

IRS's Tax Gap Estimate Remains Stable, but IRS Has Not Examined Root Causes of Undetected Noncompliance

The Tax Gap Has Remained 3.5 Percent or Less of Gross Domestic Product over 20 Years

IRS has been using NRP data to estimate the individual underreporting tax gap, starting with data from tax year 2001. The nominal value of the gross tax gap estimates has almost doubled in size, from $345 billion in tax year 2001 to a projected $688 billion for tax year 2021. However, the voluntary compliance rate has remained relatively stable ranging from 82 percent to 85 percent.[15] Likewise, the tax gap, as measured relative to the overall U.S. economy, has remained relatively stable.

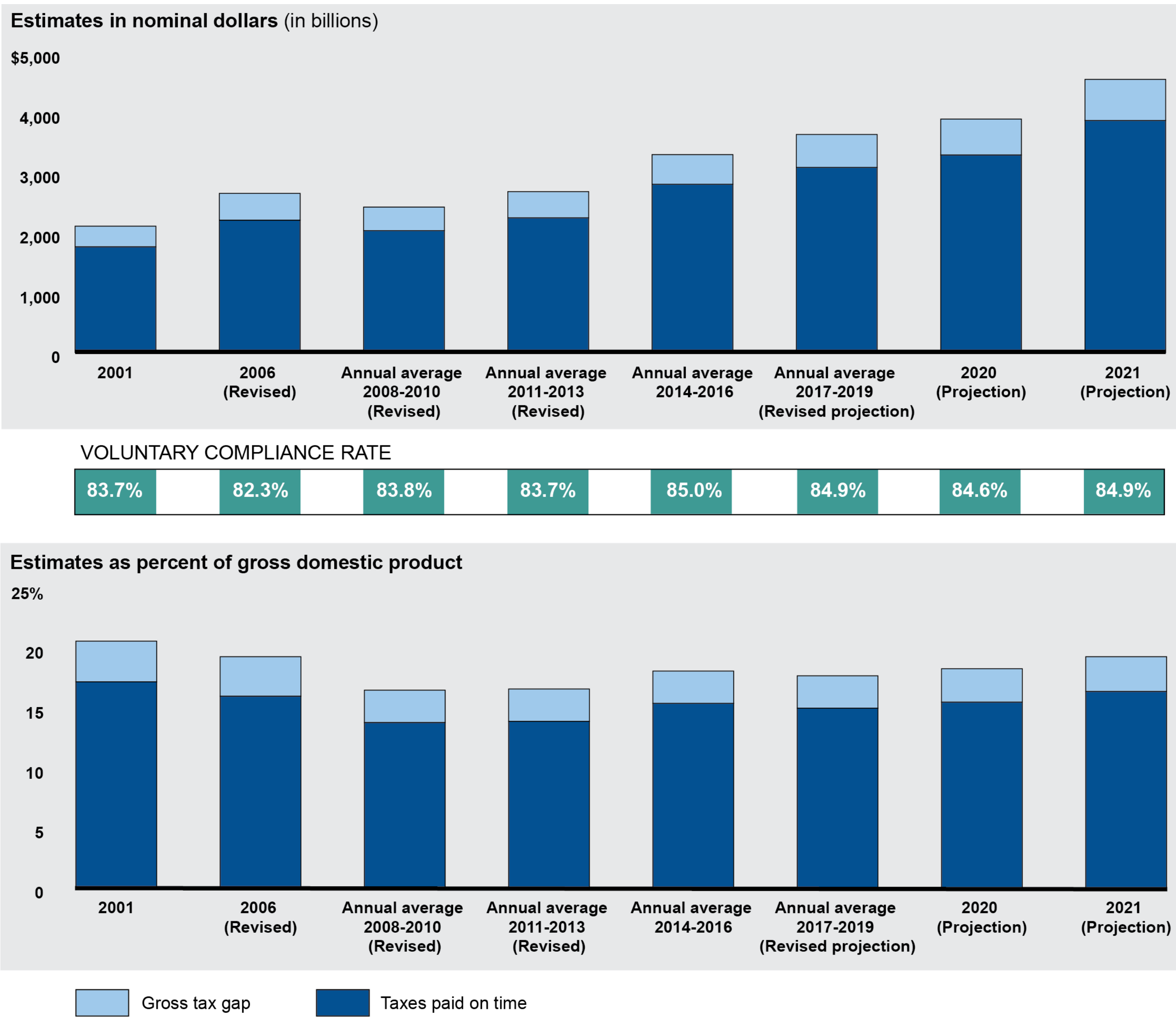
To illustrate the size of the tax gap relative to the overall economy, we present IRS's estimates of the total tax liability broken down by the gross tax gap and the amount of taxes voluntarily paid as a percentage of gross domestic product (GDP).[16] Figure 5 shows the gross tax gap has fluctuated between 2.6 percent to 3.5 percent of GDP from 2001 to 2021.[17]

[15]The voluntary compliance rate is a measure of relative compliance and defined as the amount of taxes paid voluntarily and on time divided by the tax liability, expressed as a percentage.

[16]GDP is the value of all final goods and services produced within the borders of a country in a given period. It is generally used as a measure of the size of a country's economy.

[17]We normalized IRS's tax gap estimates as a percentage of GDP to account for both inflation and real growth in the economy.

Figure 5: IRS Estimates of Gross Tax Gap and Taxes Paid Voluntarily and Timely in Dollars and as a Percent of Gross Domestic Product (Tax Years 2001-2021)



Source: GAO analysis of Internal Revenue Service (IRS) data. | GAO-24-106449

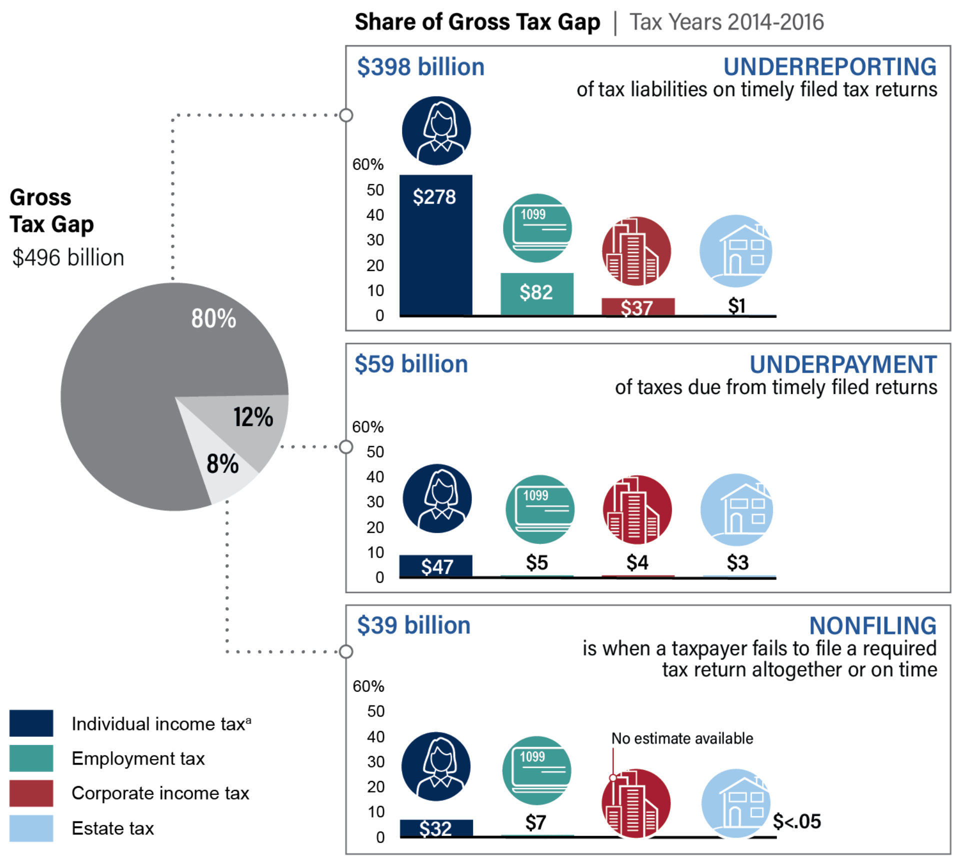
Many factors can contribute to differences in tax gap estimates over time, such as the overall level of economic activity, changes in tax law, updated data and improved methodologies, and changes in the underlying compliance behavior of taxpayers, tax preparers and other tax advisors.

Individual Underreporting Accounts for the Largest Share of the Tax Gap

As in prior estimates, individual underreporting continues to make up the largest share of the tax gap. According to IRS's tax years 2014-2016 estimate, 80 percent of the gross tax gap—about $398 billion—is a result of income underreporting by individuals and pass-through businesses such as partnerships that allocate their income or losses to their taxable partners including individuals. Specifically:

- Individual underreporting accounts for $278 billion—56 percent—of the gross tax gap.
- Underreporting of employment tax accounts for $82 billion—17 percent—of the gross tax gap.
- Underpayment of taxes accounts for about $59 billion—12 percent—while nonfiling of taxes accounts for about $39 billion—8 percent—of the gross tax gap (see fig. 6).

Figure 6: Estimated Average Annual Gross Tax Gap by Type of Noncompliance and Tax, Tax Years 2014-2016



Source: GAO analysis of Internal Revenue Service (IRS) data. GAO illustrations. | GAO-24-106449

[a]Individual income tax can include taxes from certain types of businesses, such as sole proprietorships, partnerships, or S-corporations.

Income reported by individuals includes income from certain types of businesses, such as sole proprietorships, partnerships, and S-corporations. These types of income make up most of the individual underreporting tax gap:

- Sole proprietor income accounts for the largest share (an estimated $80 billion or 29 percent) of the individual underreporting tax gap for tax years 2014-2016.
- Income from partnerships, S-corporations, estates, and trusts makes up $25 billion or 9 percent of the individual underreporting tax gap.

- Income from rents and royalties accounts for $21 billion or 7 percent of the individual underreporting tax gap (see table 1).

Table 1: Estimated Average Annual Individual Income Underreporting Tax Gap by Tax Return Item or Category, Tax Years 2014-2016

Tax return items or category	Tax gap estimate (dollars in billions)
Business income	130
Sole proprietor income	80
Income from partnerships, S-corporations, and estates/trusts	25
Rents and royalties	21
Farming income/loss	5
Non-business income	60
Other income	16
Capital gains	18
Taxable Social Security benefits	6
Pension and annuities	7
Wages	7
Sale of business property	4
Dividends	1
Interest	1
State income tax refund	1
Unemployment compensation	[a]
Credits	42
Earned Income Tax Credit	28
Child Tax Credit and Additional Child Tax Credit	8
Education credits	5
Other credits	2
Deductions	22
Unallocated marginal effects[b]	11
Exemptions	8
Filing status	5
Other taxes	4
Adjustments	-5
Total individual income underreporting	**278**

Legend: Shaded rows are subtotals.

Source: GAO analysis of Internal Revenue Service tax gap data. | GAO-24-106449

Note: Total may not add due to rounding.

[a]Less than $0.5 billion.

[b]Unallocated marginal effects reflect the difference between (1) the estimate of the individual income tax underreporting tax gap where underreported tax is calculated based on all misreporting combined, and (2) the estimate of the individual income tax underreporting tax gap based on the sum of the tax gaps associated with each line item where the line-item tax gap is calculated based on the misreporting of that item only. There may be a difference whenever more than one line item has been misreported on the same return and the combined misreporting results in a higher marginal tax rate than when the tax on the misreported amounts is calculated separately.

We analyzed NRP sole proprietor income data, since they are the single largest contributors to the tax gap estimates.[18] The following analyses are based on noncompliance that was detected in NRP audits on average for tax years 2014-2015.[19] The analyses do not account for the statistical adjustments IRS makes when estimating the tax gap for unreported income that examiners do not detect during an audit as discussed below. Therefore, these data are not readily comparable to the tax gap data presented elsewhere in this report, but they do provide insights on taxpayer behavior.

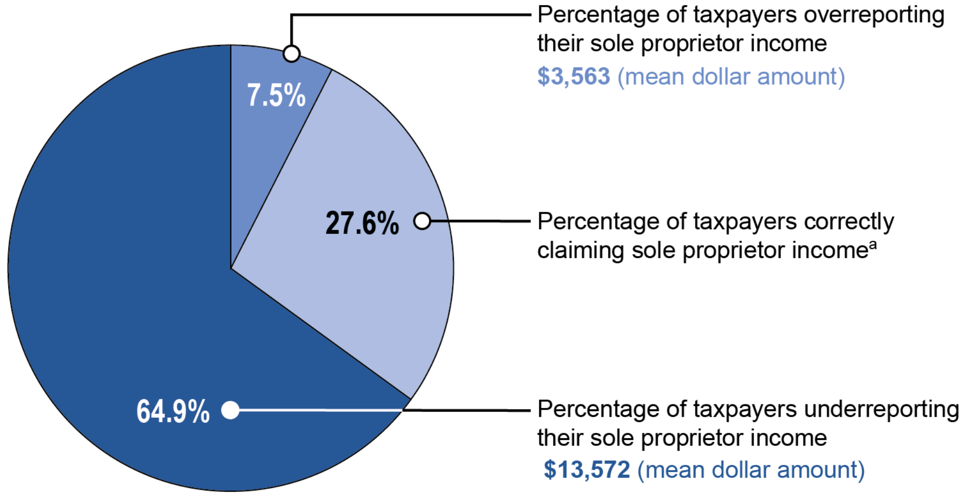
According to our analysis, on average for tax years 2014-2015, an estimated 21.5 million taxpayers reported some sole proprietor income. Of these taxpayers, an estimated 28 percent correctly reported the income. However, about 65 percent of taxpayers underreported their sole proprietor income by an average of about $13,500. The remaining 7.5 percent of taxpayers overreported their income by an average of about $3,500 (see fig. 7).[20]

[18]We have also recently reported on ways IRS could increase sole proprietor compliance and provided 17 options for reducing the tax gap for sole proprietors, including improving education and leveraging data. See GAO, *Sole Proprietor Compliance: Treasury and IRS Have Opportunities to Reduce the Tax Gap*, GAO-24-105281 (Washington, D.C.: Oct. 19, 2023).

[19]We analyzed NRP data for tax years 2014 and 2015 because they were the two most recent comprehensive NRP samples. Due to resource constraints, the 2016 NRP sample focused the audit efforts on returns that claimed certain tax credits.

[20]Overreported income would generally cause taxpayers to overstate their true tax liability. The gross tax gap is calculated as the amount of understated tax minus the amount of overstated tax.

Figure 7: Estimated Misreporting of Sole Proprietor Income, Annual Average Tax Years 2014-2015



Source: GAO analysis of Internal Revenue Service (IRS) data. | GAO-24-106449

Note: All percentage estimates in this figure have a margin of error of plus or minus 0.2 percentage points or less for a 95 percent confidence interval. All estimated mean dollar amounts have a relative margin of error of plus or minus 28 percentage points or less for a 95 percent confidence interval. Estimated percentages and mean amounts are based on noncompliance that examiners detected in National Research Program audits and do not account for undetected noncompliance in the tax gap.

[a]Amount correctly reported does not have a dollar amount associated with it because there is no difference between taxpayer reported income and income after an IRS audit.

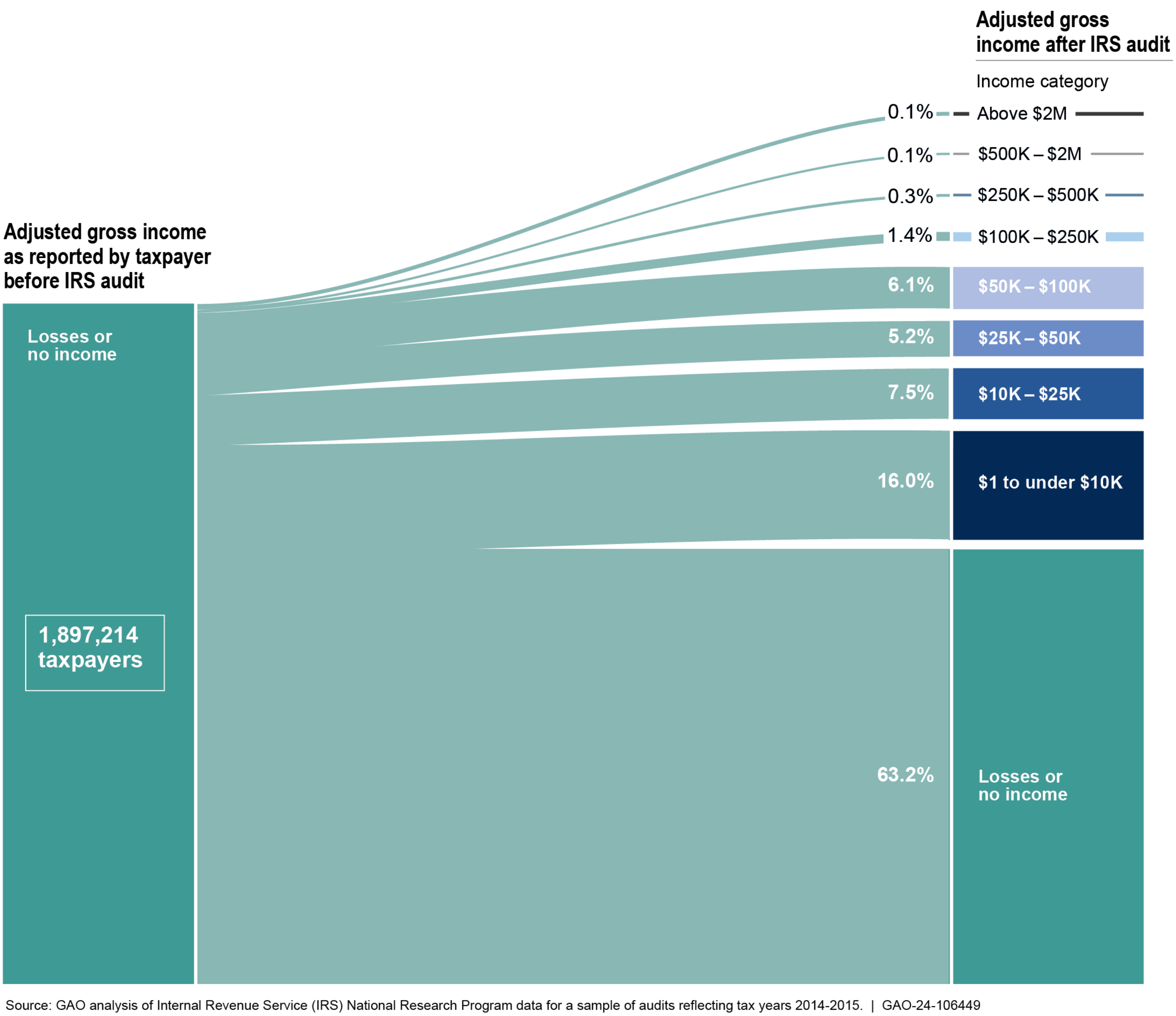
Stakeholders we spoke with expressed an interest in understanding how the tax gap is distributed by taxpayer income. However, identifying noncompliance by income is challenging because, by definition, noncompliant taxpayers do not accurately report their income. Thus, IRS does not know a taxpayer's true income without first conducting an audit.

To illustrate this challenge, we looked at how taxpayers' income changed after an audit. We analyzed the adjusted gross income as reported by the taxpayer on the tax return and compared that to the adjusted gross income as determined by IRS examiners.[21] Because a taxpayer may overstate or understate income, the taxpayer's actual income determined

[21]Adjusted gross income is the sum of income from all sources minus certain adjustments such as educator expenses, student loan interest, alimony payments, and retirement contributions. We chose this variable because a taxpayer's tax liability is based on this number. When classifying income for audit selection, IRS generally uses total positive income, which excludes losses.

from the audit may move the taxpayer from one income category to another. Looking at taxpayers reporting losses or no income reveals that an estimated 63 percent of taxpayers stayed in that category, but about 37 percent of the estimated 1.9 million taxpayers in the category moved into a higher income category—sometimes substantially higher (see fig. 8).

Figure 8: Income before and after IRS Audit for Taxpayers with Losses or No Income, Annual Average Tax Years 2014-2015



Source: GAO analysis of Internal Revenue Service (IRS) National Research Program data for a sample of audits reflecting tax years 2014-2015. | GAO-24-106449

Note: Percentages may not sum to 100 percent due to rounding. Estimated counts and percentages are based on noncompliance that examiners detected in National Research Program audits and do not account for undetected noncompliance in the tax gap. All percentage estimates in this figure have a margin of error of plus or minus 0.2 percentage points or less for a 95 percent confidence interval. The estimated number of taxpayers is equal to the sum of estimated taxpayers for all after-audit income categories. The estimated number of taxpayers who reported losses or no income has a relative margin of error of 23 percentage points for a 95 percent confidence interval.

Taxpayers across all income brackets see changes in their income because of IRS audits. In the aggregate, more taxpayers underreport income than overreport it. Therefore, when tax gap assessments are based on examiner-adjusted income rather than taxpayer-reported income, taxpayers tend to shift into higher income categories. This suggests that IRS does not know a taxpayer's true income until after an audit. More granular analyses can help provide additional insights into taxpayer behavior and could be used to develop compliance strategies.

An IRS official told us they planned to present additional analysis of the tax gap, including income distributions, in future tax gap reports. Research, Applied Analytics and Statistics (RAAS) officials recognized stakeholders' desire to have these additional data.

Nearly Half of the Individual Underreporting Tax Gap Is Based on Estimated Undetected Noncompliance; IRS Has Not Examined Its Root Causes

IRS historical research found that not all underreported income was detected in every audit because examiners vary in their ability to find income. To address this issue, IRS applies a statistical technique to the results of NRP audits. This technique—called detection controlled estimation (DCE)—is used to estimate the noncompliance that would have been detected by a hypothetical "best" examiner.[22] IRS officials then increase the detected noncompliance by an estimate of what would have been detected by the hypothetical best examiner.[23]

IRS has training, quality control, and review processes to help account for differences in examiner skills and experience. However, IRS has not taken steps to understand the specific causes of undetected noncompliance, including what may or may not be reflected in the DCE adjustments. External stakeholders we interviewed had concerns about the accuracy and reliability of DCE. Several stakeholders cited a paper by researchers at the Tax Policy Center that identifies potential reasons why DCE may either overestimate or underestimate true noncompliance.[24]

[22]To be able to estimate the effects of different examiners, the technique generally requires each examiner audit 15 or more returns with the same income line items. For example, one examiner would need to audit 15 returns each with Schedule C income in order to estimate undetected noncompliance associated with Schedule C income by that examiner.

[23]Audit results are adjusted by line item, or groups of line items, based on the relative outcomes of different examiners, controlling for other factors.

[24]Daniel Hemel, Janet Holtzblatt, and Steve Rosenthal, "The Tax Gap's Many Shades of Gray," Tax Policy Center (February 2022).

IRS estimates that nearly half of all individual underreporting, on average, is not detected by Small Business/Self-Employed (SB/SE) examiners who conduct the audits. Before the DCE adjustment is applied, the individual underreporting tax gap was $145 billion for tax years 2014-2016. After the DCE adjustment is applied, that amount increases to $278 billion for tax years 2014-2016. In at least one previous tax gap estimate, the relative DCE adjustment was even larger.

Some components of the tax gap have higher estimated undetected noncompliance than others. If these estimates are accurate, these differences suggest that the benefits of reducing undetected noncompliance may be larger for some components than others, depending on the causes and costs of doing so. For example, the tax gap estimate for sole proprietor income in tax years 2014-2016 was estimated to be $34 billion before the DCE adjustments were applied. After the DCE adjustments, the tax gap estimate for sole proprietor income was $80 billion, an increase of over 135 percent over the original estimate. In comparison, overall individual underreporting increased by about 92 percent after the DCE adjustments were applied. Appendix II provides more details on DCE adjustments and the 2014-2016 tax gap estimates.

RAAS officials acknowledged the uncertainties surrounding their estimates of undetected noncompliance and the DCE methodology and said it is unclear if DCE overestimates or underestimates noncompliance. For example, DCE may underestimate noncompliance if certain types of noncompliance go undetected by all examiners. On the other hand, if examiners incorrectly identify noncompliance, DCE may adjust examiners' results to align with the incorrectly identified noncompliance. Further, officials noted uncertainty could increase because NRP sample sizes were reduced. IRS officials stated that the smaller sample sizes for future tax gap estimates—about 4,000 returns per year—may present challenges to DCE (see more on sample size reduction below).[25]

Though DCE suggests significant noncompliance goes undetected by SB/SE examiners, IRS's tax gap analysis does not attempt to identify the root causes, including the possibility that a portion may be misestimated due to limitations with the methodology. RAAS officials told us they do not seek or provide feedback to business operating divisions about where

[25]A smaller sample size may make it harder for IRS to obtain the audit results that are needed for the DCE model, for example, since one examiner may not audit 15 returns with the same line items. IRS officials told us they may have to supplement the data with data from operational audits or historical NRP data because of reduced sample sizes.

DCE may be identifying undetected noncompliance or, for example, how examiner training may be improved based on DCE results. In addition, SB/SE officials, whose examiners conduct the audits, were unfamiliar with DCE.

Standards for Internal Control in the Federal Government states that management should use quality information to achieve the entity's objectives.[26] Further, it calls for management to communicate that information internally and externally.[27] Additionally, the Office of Management and Budget Information Quality Guidelines describe how information quality, including objectivity, are integral to every step of developing information an agency disseminates.[28] The guidelines state that objectivity is achieved by using reliable data sources and sound statistical techniques.

Given the magnitude of the estimate of undetected noncompliance, it is important for both IRS and external stakeholders to understand and have confidence in the DCE adjustment and the examiners' ability to detect noncompliance. RAAS officials told us it is difficult to validate the data through field testing because IRS cannot audit the same taxpayer twice. While such a validation strategy may not be feasible, any efforts by IRS to analyze and explain the DCE results—such as conducting descriptive or root cause analyses—could be beneficial. For example, IRS officials told us that they may be able to analyze the DCE results to see if undetected noncompliance could be explained by examiner characteristics, such as years of experience or pay grade. They may also be able to analyze whether estimated noncompliance is consistent across examiners, or different line items, over time.

Such analyses could provide a better understanding of the DCE adjustment results and could improve stakeholder confidence in the accuracy and reliability of the tax gap estimates. It could also provide potentially actionable insights to IRS officials and policymakers. For example, IRS could use the results to identify areas where examiners, as a whole, are more likely to miss noncompliance. By strengthening training

[26]GAO-14-704G, para 13.01.

[27]GAO-14-704G, paras 14.01, 15.01.

[28]Office of Management and Budget, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by Federal Agencies*; *Republication*. 67 Fed. Reg. 8452, 8453 (Feb. 22, 2002).

and procedures related to these issues, IRS could potentially improve the quality of audits.

IRS officials told us they developed pilots to provide more information on the issue. In one such pilot, examiners from two different IRS groups examined a sample of individual taxpayer returns with total positive income greater than $30 million. The returns were independently classified by IRS's Large Business and International Division's Global High Wealth examiners, in addition to the SB/SE examiners who conduct the NRP examinations. IRS officials said the pilot's initial results showed that the Global High Wealth examiners generally identified different potential compliance issues than the SB/SE examiners. If these issues are generally not detected by SB/SE examiners, they also would not be reflected in IRS's estimates of undetected noncompliance.

While IRS's pilot is a positive first step, additional analyses of the DCE technique and results could provide better insights into undetected noncompliance and address concerns over the methodology. Those results could also be used to strengthen examiner's training in areas that are systematically undetected. Such information may be especially important given the reduced sample size in future tax gap estimates. It may also help IRS address recommendations related to using NRP and tax gap data to develop compliance strategies and better align resources.[29]

[29]GAO, *Tax Gap: IRS Needs Specific Goals and Strategies for Improving Compliance*, GAO-18-39 (Washington, D.C.: Oct. 31, 2017); and Treasury Inspector General for Tax Administration, *Opportunities Exist for the IRS to Develop a More Coordinated Approach to Examination Workplan Development and Resource Allocation*, 2023-30-008 (Washington, D.C.: Feb. 8, 2023).

IRS Is Using Artificial Intelligence Models for NRP Audit Selection, but Does Not Have Controls Fully Aligned with Key Accountability Practices

IRS Redesigned NRP to Improve Efficiency and Insights

The NRP sample size has fluctuated over the years but, beginning in 2016, it was significantly reduced, as shown in table 2. According to IRS, this reduction is due to declining resources. In tax year 2019, IRS began a pilot using artificial intelligence (AI) techniques for sampling tax returns for NRP audits. In that year, IRS used this new methodology to select half of the 8,000 returns for audit along with an equivalently sized sample using its traditional sample selection process.[30] RAAS officials said that going forward they expect the overall sample size to remain at 4,000.

Table 2: National Research Program (NRP) Case Counts and Changes to NRP by Tax Year, 2001-2021

Tax Year	Approximate annual NRP sample size	Changes to NRP sample
2001	45,000	First NRP sample.
2006	12,800	Transitioned to a 3-year estimate, reducing annual sample size while maintaining statistical confidence. Began subsampling Earned Income Tax Credit (EITC) claimants through NRP to reduce need for separate EITC compliance studies.
2007-2014	12,500-14,200	No major changes from previous years.
2015	14,200	Expanded NRP data collection to determine need for improper payment reporting on the Net Premium Tax Credit.
2016	3,400	Sample reduced and limited to cases required to develop IRS's annual improper payments estimates as a temporary measure to reduce costs while redesigning the NRP sample.

[30]As of October 2023, data from tax year 2019 were not fully available for IRS officials to analyze for use in the tax gap estimate. IRS officials stated these data should be available for use in late 2024 or soon after.

Tax Year	Approximate annual NRP sample size	Changes to NRP sample
2017	4,000	Updated sample design in response to changes in tax law (2014 addition of Premium Tax Credit), need for additional data on other tax credits (e.g., American Opportunity Tax Credit) for improper payment reporting, and a reduction in examination staff. Developed new process to identify the minimum sample size required to meet program objectives.
2018	4,000	No changes from previous year.
2019	8,000	Selected 4,000 cases using traditional NRP sample selection process. Selected 4,000 cases using new sampling process using artificial intelligence (AI) as part of its efforts to redesign NRP. These cases were split into 2 samples: AI Sample 1 with 1,500 cases across the population and AI Sample 2 with 2,500 cases focused more on higher-risk and higher-income returns.
2020	1,500	Began exclusively using AI-informed sampling process. Only selected cases for AI Sample 1.
2021	4,000	Re-introduced second AI sample. Selected 1,500 cases for AI Sample 1 and 2,500 cases for AI Sample 2.

Source: GAO analysis of IRS data and reports. | GAO-24-106449

According to IRS, the redesigned sample should increase efficiency by reducing operational costs of measuring noncompliance, making better use of historical compliance data, and increasing the frequency of updates to the models and estimates.[31] IRS also undertook an initiative to make several operational improvements.[32] See appendix III for a fuller description of these operational changes.

IRS's traditional NRP sampling approach selects returns randomly from taxpayers with certain characteristics from their Forms 1040, *U.S. Individual Income Tax Return*. IRS sorts returns into groups and subgroups according to characteristics such as the level of taxpayer-reported income, whether the taxpayer is self-employed, and whether the taxpayer claimed a refundable tax credit.

The pilot sample selection process uses AI models to select returns. First, IRS categorizes all returns with the same high-level characteristics from the Form 1040 as the traditional approach. Second, the AI models assign returns a score based on predictions of the risk of noncompliance and

[31]According to IRS, the three primary uses of NRP data are the tax gap estimates, improper payment estimates, and updating audit selection models.

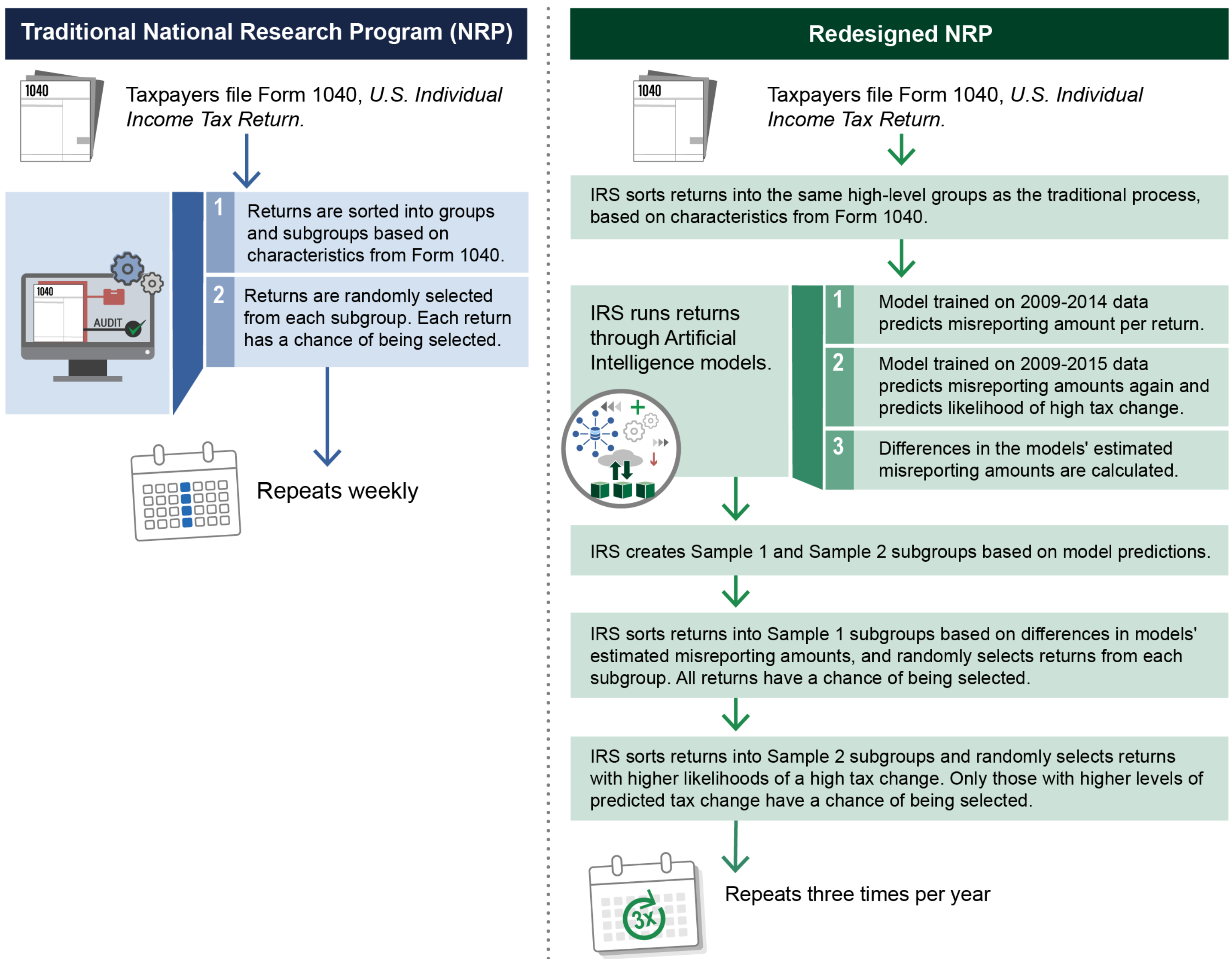
[32]We assessed IRS's sample selection redesign efforts and did not assess IRS's efforts to meet these other goals.

amount of potential tax change. Third, IRS further divides returns to draw two samples (see fig. 9).

- For AI Sample 1, IRS creates subgroups based on the degree to which cases might provide information on types of noncompliance that are new or have not been encountered recently in NRP, thereby helping to improve future model predictions of noncompliance. IRS then selects returns randomly from each subgroup. All returns in AI Sample 1 have a chance of being selected, to estimate noncompliance across the population more broadly. This is similar to the traditional sample selection process but may help to sample more returns with emerging areas of noncompliance. For tax years 2019-2021, AI Sample 1 consisted of 1,500 returns annually.
- For AI Sample 2, IRS re-sorts the remaining returns not selected in the first sample based on the likelihood of a high tax change. IRS then selects returns with a greater likelihood of a large tax change after an audit. This approach aligns with Inflation Reduction Act Strategic Operating Plan goals of auditing higher-income taxpayers and those taxpayers with higher expected levels of noncompliance. In tax years 2019 and 2021, AI Sample 2 included 2,500 returns. IRS did not select this sample in 2020.[33]

[33]A backlog of work related to the COVID-19 pandemic led IRS officials to reduce the size of the NRP sample for tax year 2020, to allow them to return to a timelier NRP schedule in future years, according to officials.

Figure 9: Traditional and Redesigned National Research Program (NRP) Sample Selection Process



Source: GAO analysis of Internal Revenue Service documents. GAO illustrations. | GAO-24-106449

IRS officials told us that despite AI methods allowing for better targeting of smaller NRP samples, the reduction in the NRP sample size will increase the uncertainty in the overall tax gap estimate. They told us smaller samples sizes may also pose challenges for estimating noncompliance for different types of income such as those shown in table

1 of this report. For example, NRP historically included more than 2,500 returns that claimed the Earned Income Tax Credit, but the redesigned sample has included less than 500 of these returns annually.

To address this, IRS reported it is exploring ways to combine operational audit data with NRP audit data when developing its tax gap estimates. IRS officials also told us that if IRS can reliably combine these data for tax gap analysis, IRS might be better positioned to identify emerging trends in noncompliance and reduce the uncertainty of the estimates due to the small sample sizes.

RAAS officials told us that the smaller NRP sample sizes and effects of the redesigned NRP selection process may lead to better information in some areas of interest such as higher-income noncompliance. However, the redesigned process may lead to less information about some specific line-items, thereby necessitating that IRS supplement NRP data with operational audit data, such as for taxpayers claiming the net premium tax credit.[34]

IRS Could Better Realize Efficiency Gains by Fully Documenting and Assessing the Redesigned NRP Sample Selection Process

We found that documentation related to the NRP redesign was limited and, in some cases, incomplete. Additionally, IRS has not developed plans to evaluate the redesigned process. According to IRS, the redesigned NRP sample is designed to be a cost-effective way to develop estimates compared to its traditional sampling process. IRS could take steps to realize these benefits by finalizing documents that outline key processes to ensure the models are implemented as intended and monitored to assess performance as IRS finalizes the pilot.

IRS Has Not Fully Documented Its AI Sample Selection Models

IRS has multiple documents that collectively provide technical details and justifications for the design of the AI models. However, no set of documents contains complete information an IRS analyst could use to run or update the models, and several key documents are in draft form. Such documentation is important to help ensure transparency and consistency in implementation.

IRS requires that program officials update internal guidance annually with procedural, organizational, or operational changes to ensure employees have the information they need to carry out their duties.[35] In addition, our

[34]According to RAAS officials, having less NRP data on certain tax credits is likely to negatively affect the timeliness of IRS's improper payment estimates.

[35]Internal Revenue Manual, § 1.11.2.2.

AI Accountability Framework (the AI Framework) outlines multiple areas in which agencies should establish structures and processes to manage, operate, and oversee the implementation of AI.[36] We found several areas in which IRS could better document its AI models to more closely align with the following key practices from the AI Framework (see table 3).

Table 3: Internal Revenue Service (IRS) Could Better Document Its Efforts to Implement National Research Program (NRP) Artificial Intelligence (AI) Models

AI Framework Key Practice	IRS Implementation Efforts
Documenting key components. Document components of AI models, including their purposes, specifications, and requirements.	We found that the draft documents included proposals for ways IRS could design its AI models and justifications for choosing those options. However, officials did not document which options and tradeoffs they ultimately selected. One draft technical document included many details that would be needed for IRS to update and validate the models in a manner that is consistent with the original model development. However, it does not describe key components, such as how IRS determined risk levels the models use to categorize returns, and when such an analysis should be updated. The document also contained some information that conflicted with other documents we reviewed.
Technical specifications. Document technical specifications to ensure the system meets its intended purpose.	IRS's finalized technical documents do not contain the level of detail that would be needed for an analyst to independently run the models or ensure consistency with past years' selection processes. For example, the documents state that data are to be divided or split before running each group of data through the model, but the documents do not specify how the data should be divided or split. IRS has another technical document, in draft form, with additional details that could help address some of the issues. However, without finalizing complete and accurate documentation outlining how the models are to be operated, IRS risks the models running inconsistently and not meeting their intended purpose.
Regularly updating AI models. Monitor AI models to determine when changes in the operational environment may affect model performance and necessitate model updates.[a]	IRS has documentation of its last update of the AI models in 2023. However, IRS documents did not outline how and when IRS officials planned to update the AI models on an ongoing basis. The documents also did not discuss how or when to update the model's inputs.
Assessing model performance. Document methods for assessing the AI system.	Conducting assessments ensures models function as intended and are sufficiently robust. Officials told us they assessed, or validated, the models when they were first developed, but IRS has not finalized documentation for how to go about this process for future updates. Without a documented plan to assess the various technical aspects of the models, IRS risks being unable to ensure consistency and comparison against past years.

Source: GAO analysis of IRS documents, interviews with IRS officials, and GAO's AI Accountability Framework. | GAO-24-106449

[a]This key practice addresses drift, establishing the range of data and model drift that is acceptable to ensure the AI system produces desired results.

Officials told us they had not completed documentation related to the NRP AI models because they consider the redesign effort to be a pilot.

[36]GAO-21-519SP.

However, IRS discontinued the traditional sampling methodology after tax year 2019. Further, IRS has decided to move forward with the AI sampling methodology for current and future NRP samples. Officials told us that they have not documented the processes for NRP AI sample selection, in part, due to disruptions related to redesigning NRP and the COVID-19 pandemic. Officials also stated they first want to assess the results of the tax year 2019 samples before completing the pilot's documentation.

During our audit, we identified two additional areas for IRS to improve its documentation in line with key practices from the AI Framework. First, we found that IRS had not defined and documented clear roles, responsibilities, and delegation of authority for carrying out each step in the redesigned sample selection process. Second, we found that IRS had not documented the variables used to develop and run the AI selection models. While our report was at IRS for comment in April 2024, IRS took steps to enhance its documentation in these two areas. These steps included drafting a memorandum listing individuals responsible for steps in the AI model development and sample selection processes. IRS also updated the finalized technical document to include details on variables and computer code used to run the AI models. These actions will increase IRS's ability to effectively implement and ensure operational effectiveness of the AI models.

Implementing a redesigned sample selection process can be a complex undertaking, especially when using emerging technologies such as artificial intelligence. The AI Framework emphasizes the importance of documentation to help ensure that the AI system's objectives are met.[37] It further emphasizes that documentation can offer a way for agencies to provide transparency, such as (1) what the system is for, (2) what it is not for, (3) how it was designed, and (4) what its limitations are.[38]

[37]GAO-21-519SP.

[38]In November 2023, the President issued an Executive Order on the development and use of artificial intelligence. Exec. Order No. 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, 88 Fed. Reg. 75191 (Nov. 1, 2023). The order describes actions to ensure responsible government use of AI, among other things. These actions include issuing guidance for agencies' use of AI, helping agencies acquire specified AI products and services, and accelerating the rapid hiring of AI professionals. We did not use this executive order as criteria as it was issued after the majority of our audit work was completed.

AI systems pose unique challenges in promoting accountability because their inputs and operations are not visible to the user. Documentation helps to promote transparency. Without complete documentation of the development and use of the NRP AI models, IRS lacks reasonable assurance that it is implementing the AI models consistently and that the models are working appropriately. Documentation will help to retain organizational knowledge when staff turnover occurs and make the methodology more transparent to future users. It would also ensure that updates continue to align with the goals of the redesigned selection process and adjust to changes in the environment.

IRS Has Not Developed an Evaluation Plan for the AI Sample Selection Process

IRS has not developed a plan for how it will evaluate the AI sample selection process, including factors to consider in deciding whether the redesigned process is achieving its intended outcomes.[39] As previously noted, IRS decided to move forward with implementing the new AI sample selection process before fully evaluating it. RAAS officials said preliminary data are showing that the new sample selection process is performing as intended and that IRS plans to evaluate the new process once tax year 2019 audit data are fully available. Officials said they expect these data will be complete in late 2024 or soon after. Nonetheless, documenting a plan to evaluate the success of the new process is key to ensuring NRP achieves its goals of providing efficiency gains over the traditional sample selection process.

As described above, IRS has a broad goal for the redesigned sample selection process to increase efficiency. Officials told us that they plan to measure the success of the redesigned process by monitoring the no-change rate of the audits, average tax adjustment by sample group, and audit case outcomes by NRP sample across the three selection processes used for tax year 2019 returns—the traditional method, AI Sample 1, and AI Sample 2.[40] IRS is tracking and documenting this information currently.

However, IRS has not developed performance targets for these metrics to serve as specific and measurable benchmarks for decision-making. Performance targets would identify a level of performance IRS would expect to see from the traditional and AI-informed samples based on past

[39]Officials told us IRS is maintaining the capability to re-implement the traditional sample selection process until it evaluates the AI pilot, although it does not plan to use the traditional sample selection process.

[40]The no-change rate is the rate at which audits close with no changes to taxes owed.

years' audit results and model testing. Because of statistical variability, it could take years for IRS to know whether the redesigned sampling strategy is meeting its objectives, including the ability of the models to correctly prioritize risk.

Standards for Internal Control in the Federal Government states that documentation is necessary to ensure a program operates effectively and consistently.[41] In addition, leading practices for pilot design state that agencies should develop a detailed data-analysis plan to track the pilot program's implementation and performance, and evaluate the results of the project and draw conclusions on whether, how, and when to integrate pilot activities into overall efforts.[42]

RAAS officials stated that developing an evaluation plan is desirable but the office responsible for this work has many competing priorities. They also said they had not yet committed to an evaluation framework because they want to be able to consider new information and context that may have been unforeseeable at the start of a pilot.

Without a documented plan for evaluating the NRP redesign effort, such as establishing performance targets, IRS lacks reasonable assurance that the redesigned process has improved efficiency. Having documented performance expectations will help IRS determine if it should continue using the new processes, revert to the previous process, or modify the process in another way to improve efficiency.

[41]GAO-14-704G, para OV4.08.

[42]GAO, *Data Act: Section 5 Pilot Design Issues Need to Be Addressed to Meet Goal of Reducing Recipient Reporting Burden*, GAO-16-438 (Washington, D.C.: Apr. 19, 2016).

IRS Seeks to Identify Emerging Issues, but Could Do More to Link Tax Gap Data to Its Strategic Operating Plan

IRS Uses Multiple Approaches to Gain Insight into Emerging Noncompliance Issues

When developing the tax gap estimate, IRS accounts for changes that occurred in the tax environment during the period covered. Changes include the introduction of new laws and regulations and any corresponding changes to tax forms, shifts in policy, and the emergence of new issues, such as increased use of digital assets and growth in online marketplaces and payment settlement entities. Identifying emerging issues can be particularly challenging for IRS because audits that identify noncompliance can take several years to complete.

To better understand these changes and related noncompliance, IRS uses information from multiple sources, including NRP and other more targeted compliance studies and research initiatives. For example:

- RAAS officials reported IRS had updated or completed several studies related to compliance within the last year and has multiple ongoing and planned initiatives that are expected to inform tax gap estimation.[43] According to RAAS officials, the agency is currently studying filing and payment compliance, EITC, unreported tip income, and offshore and large corporate compliance.
- RAAS officials reported that they are also working on multiple technical projects including some related to refining or developing alternatives to DCE and statistically combining random NRP audits with nonrandom operational audits.
- IRS officials said that their work to combine samples of NRP and samples of operational audit data may enhance their ability to identify

[43]In the past, according to the Treasury Inspector General for Tax Administration, IRS has had to curtail some of its research efforts due to resource constraints. According to officials, over time, the business operating divisions have experienced declining resources, which has affected RAAS's ability to conduct research studies. See, Treasury Inspector General for Tax Administration, *Actions Are Needed to Improve the Completeness, Development and Review of IRS Tax Gap Estimates,* 2023-10-016 (Washington, D.C.: March 2023).

emerging issues sooner, as well as mitigate some challenges associated with smaller NRP samples. RAAS officials reported that combining data from these two sources will allow them to gain earlier insights from emerging enforcement efforts being used in operational audits.

- IRS has also been able to identify emerging issues more quickly by using information from early reviews of returns from the business operating divisions to determine if returns merit audits. According to RAAS officials, this approach recently identified a new type of compliance risk associated with the 199A Qualified Business Income Deduction.[44]

IRS is also planning to undertake additional steps in response to funding received from the Inflation Reduction Act (IRA). IRS's IRA Strategic Operating Plan (SOP) describes how IRS plans to pursue enforcement for complex, high-risk, and emerging issues. To support this initiative, the IRA SOP states that IRS will improve feedback mechanisms throughout the agency to help detect emerging issues. IRS reported it plans to improve its ability to respond quickly to emerging compliance issues and develop treatments that it can deploy rapidly and integrate into existing enforcement efforts.

In February 2024, IRS released a white paper discussing the tax gap and potential revenue effects of IRA SOP initiatives aimed at improved service, providing greater transparency to taxpayers, and increasing efficiency through investment in information technology and data analytics.[45] The paper provided information on how increased revenue may accrue. For example, it explains how IRS's Information Reporting Modernization program could help increase voluntary compliance related to digital assets, as new information reporting requirements are implemented and these information returns are processed. This is based on tax gap analysis of the connection between information reporting and voluntary compliance. However, many of the efforts identified in the IRA SOP and discussed in the white paper are new or in the planning stage.

[44]This deduction allows eligible taxpayers to deduct up to 20 percent of their qualified business income, plus 20 percent of qualified real estate investment trust dividends and qualified publicly traded partnership income.

[45]IRS, *Return on Investment: Re-Examining Revenue Estimates for IRS Funding*, Publication 5901 (Washington, D.C.: February 2024).

The white paper also notes that further study of the potential benefits from the kinds of investments IRS describes in the IRA SOP is needed.

External Stakeholders Identified Options to Make the Tax Gap Estimates More Useful, but Improvements Present Tradeoffs

The tax gap estimates are used by external entities for a variety of reasons. The Taxpayer Advocate Service and Internal Revenue Service Advisory Committee reported using the tax gap estimates when considering recommendations to IRS.

Staff from some of the congressional committees we spoke to told us that they use IRS's estimates when discussing proposed policies and changes to the law. Some committee staff said that the analysis of the effect of information reporting on tax compliance is particularly useful to them. This information can be used to inform legislation on information reporting requirements. For example, staff from one committee told us the tax gap estimates were helpful to them when they were considering proposals related to the reporting threshold for Form 1099-K, *Payment Card and Third Party Network Transactions*.[46]

Committee staff and external stakeholders also highlighted various areas where the estimates could be more useful and easier to understand. In recent years, IRS has taken actions, or has plans that could help address some of the suggestions from stakeholders. For example, IRS has committed to producing annual projections to address unavoidable data lags in producing tax gap estimates. IRS is also gaining some earlier insights from operational audits, as described above. Table 4 summarizes stakeholder suggestions and IRS's actions and plans to address the feedback.

[46]*See* Pub. L. No. 117-2, tit. IX, subtit. G, pt. 8, § 9674(a), 135 Stat. 4, 185 (2021), *codified at* 26 U.S.C. § 6050W(e).

Table 4: External Stakeholders' Suggestions for Improving Tax Gap Estimates

Improvement Suggestion	Internal Revenue Service (IRS) Plan or Action to Address
Granularity Some stakeholders said they would like to see more granular information in the estimates. For example, many said they would like to see information on the distribution of the tax gap across the income spectrum so they could better understand where noncompliance exists. This information would help congressional staff develop legislation to address these issues.	• IRS is exploring ways to use operational audit data to help address limitations related to tax gap audit sample size. • In the two most recent tax gap estimates, IRS has included new information about noncompliance associated with different tax credits. • IRS's Inflation Reduction Act Strategic Operating Plan includes a key project to better understand the tax gap by enhancing the timeliness, granularity, and comprehensiveness of tax gap estimates.
Timeliness Some stakeholders said that the long lag time between the tax years studied in each published report and the release of the reports makes it difficult for them to gauge the effectiveness of legislation intended to improve compliance, or to understand changes in the tax environment.	• In its October 2022 publication on the tax gap estimates for tax years 2014-2016, IRS included projections for future (tax years 2017-2019) tax years for the first time. • IRS has said it is committed to releasing updated projections annually. In October 2023, the agency published updated projections (tax years 2020 and 2021).
Comprehensiveness Some stakeholders said IRS could explain what information is not included in the estimates more clearly. In addition, some said they would be interested in seeing more certain estimates of noncompliance in areas such as among high-income taxpayers, partnerships, individuals with offshore investments, and digital assets.	• The tax years 2014-2016 estimates included an expanded discussion of data limitations. In addition, tax years 2020 and 2021 tax gap projections included a more extensive discussion of potential sources of error, including areas that are not fully covered by the methodologies. • In its response to an April 2023 letter from Representative Bill Pascrell concerning the tax gap estimates, IRS said it has initiated research to quantify the extent of international noncompliance that may not be captured within the current estimation methodology. • In the same letter, IRS said it is drafting a formalized strategy to identify, prioritize, and develop estimates for areas of noncompliance not fully addressed in the current tax gap estimates. For example, IRS is trying to quantify aspects of noncompliance associated with foreign-sourced income, pass-through income, virtual currency holdings, and offshore income.
Methodological clarity Some stakeholders said IRS could provide a clearer explanation of the methodology used to develop the estimates and how it has changed over time. Some noted concerns about the reliability and accuracy of the estimates. Some stakeholders were also concerned about detection controlled estimation and the extent to which it affects the tax gap estimates.	• IRS has made some of the documents that describe the methodology it uses to generate the estimates publicly available. • The tax gap projections for tax years 2020 and 2021 include a table showing the data and method used to develop projections for different tax gap components.
Broader context Some stakeholders said they wanted more information about effectiveness or costs of compliance initiatives, areas where legislative action is needed and the effects of legislative actions on compliance in the tax gap estimates.	• Typically, other components within the Department of the Treasury are responsible for performing analyses of the effects of tax law. For example, the Office of Tax Analysis within Treasury analyzes the effects of the existing tax law and alternative tax programs and has prepared analytical reports on economic aspects of domestic and international tax policy. • IRS has not provided additional information on plans to address this stakeholder concern.

Source: GAO analysis of IRS documents and interviews with IRS officials, congressional committee staff, and officials from other external organizations involved in tax policy and administration. | GAO-24-106449

Any improvements to the tax gap estimates are likely to come with associated costs and tradeoffs. For example:

- To increase the granularity of the tax gap estimates, IRS may need to increase the number of NRP audits within selected categories. These audits place burden on the taxpayers whose returns are audited, and the audits also draw on IRS enforcement resources.
- There are opportunity costs associated with the time and resources employees within RAAS devote to the tax gap estimates and any improvements to them. RAAS officials told us that the tax gap estimates are currently developed by fewer than 10 staff who juggle many competing priorities. Focusing on any particular improvement may take attention away from other priorities.
- IRS may need to make technological investments, such as to its AI models, to use operational data to estimate noncompliance.

RAAS officials told us that they try to balance producing granular and timely estimates against burden and cost. They said some of the modernization efforts within IRS and additional staffing within RAAS resulting from IRA funding will help them improve the quality and detail of their estimates.

Staff from one committee said they would appreciate more context around what kind of resources would be needed to reduce the gap. They noted that some external stakeholders view the tax gap as available funds that can be captured and do not always consider the costs associated with enforcement and revenue collection efforts.

IRS Is Missing an Opportunity to Better Link Recent Compliance Strategy to NRP Data

According to IRS, the purpose and intended use of the tax gap estimates is to provide IRS with periodic appraisals of the nature and extent of noncompliance for use in formulating tax administration strategies. NRP data underlie the individual underreporting segment of the tax gap. Underreporting by individuals accounts for the largest component of the estimated tax gap.

IRS's IRA SOP describes initiatives and projects IRS will use to improve enforcement, among other things. The IRA SOP includes an objective to address the tax gap through expanded enforcement focused on certain types of taxpayers. However, we identified areas where IRS has not documented a clear link between this IRA SOP objective and the agency's data and analysis of the tax gap. For example, the IRA SOP does not include any initiatives to address the largest areas of noncompliance that are identified by the tax gap estimates:

- Sole proprietor noncompliance has consistently been the largest individual contributor to the tax gap ($80 billion annual average for tax years 2014-2016), but the IRA SOP does not describe strategies or enforcement actions to address this problem.[47] We recently reported that neither the Department of the Treasury nor IRS plans to develop an overall tax gap strategy that would include options for addressing sole proprietor noncompliance.[48]
- Income from rents and royalties ($21 billion annual average for tax years 2014-2016) is another persistent area of individual underreporting that IRS does not describe strategies to target or otherwise link in the IRA SOP.[49]

In other cases, the IRA SOP objective to address the tax gap includes initiatives that are not supported by tax gap data and analysis. For example, the IRA SOP contains an objective to focus expanded enforcement on taxpayers with complex tax filings and high-dollar noncompliance to address the tax gap. It states that IRS will increase its focus on segments of taxpayers with complex issues and complex returns where audit rates are minimal, such as those related to large partnerships, large corporations, and high-income and high-wealth

[47]In October 2023, we recommended that Congress consider requiring the Secretary of the Treasury to ensure that a tax gap strategy includes improvements to sole proprietor compliance and that those improvements are incorporated into broader tax gap reduction efforts. See, GAO, *Sole Proprietor Compliance: Treasury and IRS Have Opportunities to Reduce the Tax Gap*, GAO-24-105281 (Washington, D.C.: Oct. 19, 2023).

[48]GAO-24-105281.

[49]In August 2008, we recommended that Congress consider amending the Internal Revenue Code to make all taxpayers with rental real estate activity subject to the same information reporting requirements as other taxpayers operating a trade or business. See GAO, *Tax Gap: Actions That Could Improve Rental Real Estate Reporting Compliance*, GAO-08-956 (Washington, D.C.: Aug. 28, 2008).

individuals.[50] The IRA SOP objective does not address the fact that the three areas identified for increased focus are areas where IRS lacks the data to estimate noncompliance fully and reliably. A different section of the IRA SOP notes that limited compliance coverage in areas, such as complex partnerships and some international tax issues, left IRS with insufficient data to accurately estimate the true size of the tax gap in these segments.

The IRA SOP initiatives concerning large partnerships, large corporations, and high-income and high-wealth individuals all cite low audit rates for these types of returns. However, the IRA SOP does not describe IRS plans to use data from increased audits of these taxpayers to better understand their contributions to the tax gap or to incorporate these findings into future tax gap reporting. This contrasts with the IRA SOP initiative focused on helping taxpayers understand and claim appropriate credits and deductions, which includes a project to incorporate findings of research back into tax gap reporting.

IRS officials reported that evidence outside of NRP has identified these areas as likely areas of significant undetected noncompliance. IRS officials pointed to ongoing external research with academics and said related internal research is underway. This research, combined with the current lack of data concerning these areas, motivated their inclusion in the IRA SOP. Officials stated that expanded enforcement initiatives in these areas should improve IRS's ability to estimate noncompliance related to these types of tax filings in future tax gap estimates.

The Deputy Commissioner for Services and Enforcement told us that the IRA SOP is extensive but does not include everything IRS is doing to reduce noncompliance. The Deputy Commissioner agreed that there is not a project or initiative in the IRA SOP focused on sole proprietorships, despite multiple tax gap estimates showing that this group is the largest

[50]In July 2023, we reported on the low audit rate for large partnerships and made recommendations to improve the audit selection and tracking process. IRS agreed with three of the recommendations and partially agreed with one of the recommendations. IRS has plans to address the three recommendations with which it agreed. See GAO, *Tax Enforcement: IRS Audit Processes Can Be Strengthened to Address a Growing Number of Large, Complex Partnerships*, GAO-23-106020 (Washington, D.C.: July 27, 2023). In January 2024, we reported on IRS audits of high-income/high-wealth individuals. See GAO, *Tax Compliance: Opportunities Exist to Improve IRS High-Income/High-Wealth Audits*, GAO-24-106112 (Washington, D.C.: Jan. 11, 2024).

single contributor to the tax gap. However, the Deputy Commissioner said IRS is aware that work is needed in this area.

In addition, IRS has also not clearly defined how it plans to measure the IRA SOP indicator of a "decreased tax gap relative to the gap without the resources provided by the IRA." RAAS officials told us that they do not believe the tax gap estimate is suitable for use as a performance indicator. RAAS officials have expressed this position to us in the past. For example, they noted that IRS actions alone do not determine the level of voluntary compliance. Instead, voluntary compliance is determined by the interaction of many factors, such as taxpayer behavior, tax law complexity, and IRS resources.

The officials explained that the discussion in the IRA SOP refers to a more theoretical meaning of the tax gap and not the specific estimate. The tax gap estimate provides a general understanding of noncompliance at a point in time. The IRA SOP uses the tax gap concept that more taxes are owed than paid. Therefore, enhanced funding devoted to enforcement efforts will reduce the tax gap by recovering more of what is owed. However, the IRA SOP does not highlight this nuance, which could cause confusion among decision-makers.

Using quality information, such as NRP data, to achieve the agency's objectives is a principle of *Standards for Internal Control in the Federal Government*.[51] Further, *Standards for Internal Control in the Federal Government* recognizes documentation is a necessary part of an effective system, but the level and nature of documentation vary based on the size of an entity and the complexity of the operational process. According to *Standards for Internal Control in the Federal Government*, documentation is critical for demonstrating the design, implementation, and operating effectiveness of a system.[52] Considering the size and relative importance of the tax gap, clearly linking IRS's IRA SOP and supporting implementation plans to NRP data and the tax gap estimates would be consistent with *Standards for Internal Control in the Federal Government*.

In 2017, we reported that IRS had not documented a comprehensive compliance strategy showing how it intended to analyze and use the tax gap data, particularly from the NRP, to develop or improve compliance

[51]GAO-14-704G, para 13.01.

[52]GAO-14-704G, OV4.08.

programs.[53] We recommended that IRS develop and document a strategy outlining how it would use NRP data to update compliance strategies that could help address the tax gap. IRS agreed with this recommendation.

In response to this recommendation, IRS developed a strategy that is not yet finalized and implemented to address measurement limitations of the tax gap. The document describes IRS's plans to further research and better measure areas of noncompliance that are priorities in the IRA SOP, such as high-income and high-wealth taxpayers. While the efforts described in this strategy should improve data availability and IRS's understanding of noncompliance, IRS has not yet shown how it will use these improved data and understanding to update its compliance strategies. In addition, the plan to improve the tax gap measurement does not explain how IRS will use current NRP data to develop or improve compliance programs.

When developing the IRA SOP, IRS missed an opportunity to establish a clear link between the agency's compliance strategy and NRP data. As IRS develops implementation plans for the initiatives and projects described in the IRA SOP, it has an opportunity to establish this linkage and further integrate a significant source of evidence to refine its compliance strategies. Clearly linking compliance strategies with data would help Congress and other decision-makers be assured that IRS is effectively allocating enforcement resources.

Conclusions

Taxpayers fail to pay hundreds of billions of dollars in taxes every year, and even a modest reduction in the tax gap would yield significant financial benefits. However, reducing the tax gap continues to be a challenging task for IRS and Congress.

IRS has acted in recent years to strengthen tax gap estimates. However, to increase its assurance that the individual underreporting estimates—the largest part of the tax gap—are accurate, IRS could take steps to more fully analyze and understand the root causes of undetected noncompliance. This may help IRS increase the confidence of the statistical technique and the adjustment results. This is especially important given the reduced sample size planned for future tax gap estimates. This analysis may also lead to improvements in examiner training.

[53] GAO-18-39.

IRS is implementing artificial intelligence models that have the potential to help IRS select NRP audits more efficiently. However, it is important that IRS document elements of the AI models as it moves forward to ensure that key model components, technical specifications, model updates, and model performance are clearly communicated and assessed. In addition, developing and executing a plan to evaluate the new AI sample selection process will help IRS determine if it should continue using the new process, make adjustments, or revert to the traditional process. As IRS works to leverage technology it is important for IRS to balance potential savings against reductions in the validity and reliability of the tax gap estimates.

Additionally, we previously recommended that IRS develop and document a plan outlining how it would use NRP data to update compliance strategies that could help address the tax gap. While IRS took a first step by developing a draft plan to improve its measurement of the tax gap, it has not shown how it will integrate those improvements, or current NRP data, into future compliance strategies and implementation plans. IRS has an opportunity to more fully use the data it already collects to shape compliance strategies that are responsive to the areas of greatest need. Making data-based decisions would increase the transparency and accountability of IRS's enforcement strategies and resource allocation.

Recommendations for Executive Action

We are making the following six recommendations to IRS:

The Commissioner of Internal Revenue should ensure officials in the Research, Applied Analytics and Statistics Division work with officials in the Small Business/Self-Employed and Large Business and International Divisions to develop and report on analyses that could provide stakeholders with more confidence in the methods used to estimate undetected noncompliance, and information on its root causes. (Recommendation 1)

The Commissioner of Internal Revenue should complete documentation on key components of the NRP AI models, such as which model options were considered and chosen for implementation, and how IRS determined risk levels and time frames for risk level updates. (Recommendation 2)

The Commissioner of Internal Revenue should complete documentation for the NRP AI models on the technical specifications needed to run the model, such as how the data are to be split or divided. (Recommendation 3)

The Commissioner of Internal Revenue should ensure IRS has complete documentation on how to update the NRP sample selection process in response to changes in the operating environment, including the performance of its AI models and risk level analysis. (Recommendation 4)

The Commissioner of Internal Revenue should complete documentation on processes for assessing the results of AI model updates, including validating the models. (Recommendation 5)

The Commissioner of Internal Revenue should develop and document a plan for evaluating the redesigned sample selection process, including factors IRS will consider in determining whether or how to continue using it. (Recommendation 6)

Agency Comments and Our Evaluation

We provided a draft of this report to IRS for review and comment. IRS provided written comments, which are reproduced in appendix IV. In its letter, IRS agreed with our six recommendations and described steps it plans to take in response to each recommendation.

The draft report included two additional recommendations for IRS to fully define and document roles and responsibilities for the National Research Program Artificial Intelligence models (draft report Recommendation 2) and to complete documentation on the data sources needed to run the models (draft report Recommendation 5). In April 2024, while the draft report was at the agency for comment, IRS addressed these two recommendations. Specifically, IRS provided documentation that we reviewed and determined addressed both of the recommendations. Therefore, we removed the two recommendations and modified our report accordingly.

As agreed with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution until 30 days from the report date. At that time, we will send copies to the appropriate congressional committees. We are also sending copies of the report to the Commissioner of Internal Revenue and other interested parties. In addition, the report is available at no charge on the GAO website at http://www.gao.gov.

If you or your staff have any questions about this report, please contact me at (202) 512-6806 or mctiguej@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 V.

Sincerely,

James R. McTigue, Jr.
Director, Tax Policy and Administration
Strategic Issues

Appendix I: Objectives, Scope, and Methodology

This report (1) describes the main types of noncompliance in the Internal Revenue Service's (IRS) estimates of the gross and net tax gap for tax years 2014-2016 and changes from prior estimates; (2) assesses IRS's planned changes to the National Research Program (NRP), including any changes to sample design and use of artificial intelligence (AI); and (3) assesses the extent to which the estimate helps IRS and other stakeholders gain insights into noncompliance for formulating tax administration strategies.

To describe the main types of noncompliance in IRS's estimates of the gross and net tax gap for tax years 2014-2016 and changes from prior estimates, we reviewed IRS's current tax gap estimate and supporting documentation as well as underlying data from NRP. We used this information to describe the tax gap and the types of noncompliance that were the largest contributors to the most recent estimate, tax years 2014-2016. We compared the estimates to prior estimates from tax year 2001, tax year 2006, tax years 2008-2010, tax years 2011-2013, and projections of the tax years 2017-2019, 2020, and 2021 tax gaps to identify and report on trends. In addition, we reviewed our past reports and Treasury Inspector General for Tax Administration (TIGTA) reports on IRS's previous tax gap estimates.

We analyzed NRP data that underlie IRS's tax gap estimates for individual income tax returns for combined tax years 2014 and 2015 to show additional details on noncompliance that examiners detected in NRP audits, a key component of the individual underreporting tax gap. We used data from these years because they were the two most recent comprehensive NRP samples. Due to resource constraints, the 2016 NRP sample was restricted to returns that claimed certain tax credits. We generated weighted estimates of the percentage of returns with any error, total taxpayers with any error, and mean dollar amount error from the NRP sample. All percentage estimates presented in this report have a margin of error no greater than plus or minus 0.2 percent at the 95 percent level of confidence unless otherwise noted. All taxpayer and mean dollar estimates have a relative margin of error of plus or minus 20 percent at the 95 percent confidence level unless otherwise noted. We used the data to present information on the estimated misreporting of sole proprietor income and income before and after IRS examination for taxpayers for tax years 2014-2015.

To assess the reliability of IRS's tax gap estimates, projections, and NRP data, we reviewed IRS technical papers and reports as well as papers by researchers to understand and describe the methodologies IRS used to estimate each component of the tax gap, as well as changes in those methodologies. We also conducted electronic testing of the data.

Further, we interviewed officials from IRS's Office of Research, Applied Analytics and Statistics (RAAS) who are responsible for estimating the tax gap about the methodologies used and the strengths and limitations of the data. We also interviewed RAAS officials about the detection controlled estimation (DCE) technique IRS uses to account for undetected noncompliance in its estimates. We determined these data were sufficiently reliable to discuss the component parts of the tax gap and present trends in the tax gap estimates over time. We also discuss a key limitation in the report related to DCE. We compared that technique against *Standards for Internal Control in the Federal Government* and the Office of Management and Budget Information Quality Guidelines.[1]

To assess IRS's planned changes to NRP, including any changes to sample design and use of AI, we reviewed IRS documents containing proposals for redesigning NRP and technical information about NRP's sample selection and AI models. We interviewed IRS officials about goals for the program redesign, options IRS could pursue for the redesign, and recommendations IRS working groups made for pilot efforts. Documents and officials described technical changes to the sample design and potential operational changes in NRP and IRS's Small Business/Self-Employed Division, the business unit that conducts NRP audits. We reviewed IRS tax gap reports and technical papers and TIGTA reports that described NRP processes.

We assessed IRS documents and interviewed officials to understand the use of AI as a key feature of the NRP redesign. These documents described the technical AI or machine-learning processes IRS used to sort and select taxpayers as part of the redesigned NRP. We also interviewed IRS officials responsible for the redesign efforts and the use of AI in those efforts and attended quarterly briefings in which IRS detailed progress and developments related to the redesign.[2] We also reviewed IRS information showing the results of cases selected by the different sampling methods, including those selected by traditional means

[1]GAO-14-704G and Office of Management and Budget, *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by Federal Agencies; Republication*. 67 Fed. Reg. 8452 (Feb. 22, 2002).

[2]IRS began providing quarterly briefings to us about the NRP redesign in 2019. We used information IRS provided in those briefings as part of this audit.

as well as those selected by AI. We also assessed IRS's plans to evaluate the performance of the NRP AI models.

We assessed IRS's use of AI and related plans against *Standards for Internal Control in the Federal Government* related to documentation, selected practices in our AI Accountability Framework (the AI Framework), and leading practices for pilot design.[3] We selected practices from the AI Framework against which to assess the NRP AI models based on their relevance to the NRP AI models, including practices related to documentation and evaluating the results of AI models.

To assess the extent to which the estimates help IRS and other stakeholders gain insights into noncompliance for formulating tax administration strategies, we reviewed documentation on the purpose and intended use of the tax gap estimates and how IRS identifies and accounts for emerging trends. We spoke to IRS officials from RAAS as well as the four business operating divisions to determine how the tax gap estimates are used internally and how RAAS and the operating divisions coordinate their efforts. We also spoke to the Deputy Commissioner for Services and Enforcement to determine how the tax gap estimates influence agency-wide tax administration strategies.

We conducted semi-structured interviews with staff from seven congressional committees with an interest in tax writing as well as IRS funding or federal budget oversight. We selected the Joint Committee on Taxation (JCT), the Senate Committee on Finance, the House Committee on Ways and Means, the Senate Committee on the Budget, the House Budget Committee, and the Senate and House Committees on Appropriations. We held seven meetings—a combined meeting for JCT and separate meetings for minority and majority staff from the other committees we selected. We invited key staff from these committees to participate in semi-structured interviews about how they use IRS's tax gap estimates and ways in which the estimates could be made more useful to them. We also conducted semi-structured interviews with external stakeholders from the National Taxpayer Advocate Service, Internal Revenue Service Advisory Committee, and the Electronic Tax Administration Advisory Committee about the extent to which they use the

[3]GAO-14-704G; GAO-21-519SP; and GAO-16-438.

tax gap estimates and their thoughts on the information included in the estimates.

We spoke to IRS officials and reviewed documents to understand how IRS accounts for changes in the tax environment that could affect the tax gap estimates and how, if at all, IRS is ensuring initiatives described in the Inflation Reduction Act (IRA) Strategic Operating Plan (SOP) will result in data for use in developing the tax gap estimates.

We reviewed IRS's IRA SOP to identify connections between this plan and tax gap data as presented in the most recent tax gap estimates and tax gap projections. We compared the information in these two sources to see if the findings of the most recent tax gap estimates were reflected in the agency's strategy and supplemented our review with additional information derived from our conversations with IRS officials. We assessed our findings against *Standards for Internal Control in the Federal Government r*elated to information and communication and documentation.

We conducted this performance audit from December 2022 to May 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: Additional Information on Undetected Noncompliance in Tax Gap Estimates

The Internal Revenue Service (IRS) applies a statistical technique called detection controlled estimation (DCE) to the results of National Research Program audits. This technique adjusts for estimated differences in examiners' abilities to detect noncompliance. According to IRS officials, this approach is a reasonable method for accounting for undetected noncompliance in underreported individual income and is the best currently available. As shown in table 5, almost half of the underreporting component of the tax gap consists of income that IRS did not detect during examinations. DCE is not applicable to the underpayment and nonfiling components of the tax gap.

Table 5: Estimated Average Annual Individual Income Underreporting Tax Gap by Tax Return Item or Category, Tax Years 2014-2016

Tax Return Items or Category	Tax Gap Estimates before Applying DCE Adjustment (dollars in billions)	Tax Gap Estimate after Applying DCE Adjustment (dollars in billions)	Percentage Change
Business Income	$48	$130	171%
Sole Proprietor Income	34	80	135
Income from partnerships, S-corporations, and estates/trusts	6	25	317
Rents and royalties	6	21	250
Farming income/loss	2	5	150
Nonbusiness income	21	60	186
Other income	7	16	129
Capital gains	4	18	350
Taxable Social Security benefits	2	6	200
Pension and annuities	4	7	75
Wages	3	7	133
Sale of business property	1	4	300
Dividends	[a]	1	[b]
Interest	[a]	1	[b]
State income tax refund	[a]	1	[b]
Unemployment compensation	[a]	[a]	[b]
Credits	38	42	11
Earned income tax credit	24	28	17
Child tax credit and additional child tax credit	8	8	0
Education credits	5	5	0
Other credits	2	2	0
Deductions	20	22	10

Tax Return Items or Category	Tax Gap Estimates before Applying DCE Adjustment (dollars in billions)	Tax Gap Estimate after Applying DCE Adjustment (dollars in billions)	Percentage Change
Unallocated marginal effects[c]	3	11	267
Exemptions	6	8	33
Filing Status	5	5	0
Other Taxes	3	4	33
Adjustments	-1	-5	400
Total individual income underreporting	**145**	**278**	**92**

Legend: Shaded rows are subtotals. DCE = Detection Controlled Estimation.

Source: GAO analysis of IRS Tax Gap data. | GAO-24-106449

Note: Components might not sum to totals because of rounding.

[a]Less than $0.5 billion or 0.5 percent.

[b]Unable to calculate percentage change because values are not precisely specified.

[c]Unallocated marginal effects reflect the difference between (1) the estimate of the individual income tax underreporting tax gap where underreported tax is calculated based on all misreporting combined, and (2) the estimate of the individual income tax underreporting tax gap based on the sum of the tax gaps associated with each line item where the line item tax gap is calculated based on the misreporting of that item only. There may be a difference whenever more than one line item has been misreported on the same return and the combined misreporting results in a higher marginal tax rate than when the tax on the misreported amounts is calculated separately.

Appendix III: National Research Program Redesign Operational Changes

After consulting with internal and external stakeholders, the Internal Revenue Service (IRS) developed a proposal for redesigning the National Research Program (NRP) starting with tax year 2019. The proposal included recommendations for changes to sample design and other operational changes in NRP and the Small Business/Self-Employed division to achieve five goals.[1] Table 6 lists the redesign goals and efforts aligned to each goal.

Table 6: Internal Revenue Service (IRS) Efforts to Achieve the Tax Year 2019 National Research Program (NRP) Redesign Goals

Goal	Efforts
Reduce operational costs of measuring noncompliance	• Reduce the number of NRP cases each year by improving sample selection and adjusting reporting requirements (e.g., decreasing the number of cases IRS data show are less influential for estimating IRS's risk-based operational audit selection models, and those required for improper payment estimates). • Automate some administrative processes required to prepare selected returns for audit (e.g., determination of which issues will be examined). • Provide an unbiased representation of the nature and extent of noncompliance by using both statistically adjusted data and operational audit data selected with NRP's new artificial intelligence (AI) sample selection process.
Increase the currency of models and estimates to improve decision support and operational effectiveness	• Automate some administrative processes required to prepare selected returns for audit (e.g., determination of which issues will be examined).
Increase employee and taxpayer satisfaction	• Better leverage skills and expertise of field agents by selecting cases that historical data show have a higher risk of noncompliance. • Reduce the number of NRP cases that have a lesser effect on updating IRS risk-based audit selection models, thereby reducing taxpayer burden.
Make better use of all available compliance data	• Leverage historical NRP data to train AI models to select NRP cases based on observed noncompliance. • Streamline NRP and operational audit processes so data from both NRP and operational audits can be used in the future for estimating the tax gap, improving IRS's risk-based audit selection models, and estimating improper payments.
Increase the speed of learning and improvement of risk models.	• Use AI models to select NRP cases. Specifically, design models to help select complementary samples that prioritize both known risks and areas of emerging noncompliance. As models are trained on more recent data, the sample will adapt based on new types of noncompliance that were detected.[a]

Source: GAO analysis of IRS information. | GAO-24-106449

[a]Using funds from the Inflation Reduction Act, IRS has a service-wide initiative to develop a process for continually refining compliance analytics models, including the NRP AI models, based on feedback and new information.

[1]We assessed IRS's sample selection redesign efforts and did not assess IRS's efforts to meet these other goals.

Appendix IV: Comments from the Internal Revenue Service

DEPARTMENT OF THE TREASURY
INTERNAL REVENUE SERVICE
WASHINGTON, DC 20224

CHIEF OPERATING OFFICER

April 17, 2024

Mr. James R. McTigue, Jr.
Director, Tax Policy and Administration
U.S. Government Accountability Office
441 G Street, N.W.
Washington, D.C. 20548

Dear Mr. McTigue:

Thank you for the opportunity to review your draft report titled *Tax Gap: IRS Should Take Steps to Ensure Continued Improvement in Estimates.* The IRS Office of Research, Applied Analytics and Statistics (RAAS) is the Service's centralized research and analytic organization. One of RAAS' responsibilities is the development of tax gap estimates. The tax gap is a measure of the level of overall noncompliance, in the context of Internal Revenue Code provisions in effect at the time. The estimates provide the IRS with periodic appraisals of the nature and extent of noncompliance for use in formulating tax administration strategies. The word "tax" in the phrase "tax gap" is used broadly to encompass not only tax but also refundable and nonrefundable tax credits. The gross tax gap is the difference between true tax liability for a given tax year and the amount that is paid on time. It is comprised of the non-filing gap, the underreporting gap, and the underpayment (or remittance) gap. The net tax gap is the portion of the gross tax gap that will never be recovered through enforcement or other late payments.

RAAS is committed to continuously improving the tax gap estimates. For example, in Fiscal Year (FY) 2022 RAAS released tax gap estimates covering the Tax Year (TY) 2014-2016 time frame along with the first release of tax gap projections covering the TY 2017–2019 time frame. Additionally, RAAS is committed to providing timely tax gap information through annual updates to the tax gap projections. RAAS released the first annual updated projections in FY 2023 covering TY 2020 and TY 2021. We appreciate your acknowledging that the addition of the projections to the tax gap reports, in part, addresses stakeholder concerns with the timeliness of the estimates. RAAS is continuing to research ways to improve our estimation methodology where data are limited or not readily available, including better understanding the root causes of undetected noncompliance.

We also want to thank you for highlighting the extensive documentation that already exists for the redesign of the National Research Program. Existing documentation has allowed for full replication of the risk models by junior RAAS analysts. We agree that there is always room for improving existing technical documentation to make it more accessible, and we recognize the importance of formally documenting analyst and

2

management decisions. We will incorporate your recommendations into our processes moving forward.

We agree with all proposed recommendations. As noted, we have already submitted additional documentation demonstrating that we implemented two of your recommendations; and, as a result, you have agreed to remove them from the final report. We are attaching our corrective action plan for all original recommendations. If you have any questions, please contact me, or a member of your staff may contact Peter Rose, Director, Knowledge Development and Application, at 202-803-9524.

Sincerely,

Melanie R. Krause

Digitally signed by Melanie R. Krause
Date: 2024.04.17 21:18:20 -04'00'

Melanie R. Krause
Chief Operating Officer

Enclosure

Enclosure

GAO Recommendations and IRS Responses to GAO Draft Report
Tax Gap: IRS Should Take Steps to Ensure Continued Improvement in Estimates
(GAO-24-106449)

Recommendation 1:
The Commissioner of Internal Revenue should ensure officials in the Research, Applied Analytics and Statistics (RAAS) Division work with officials in the Small Business / Self-Employed (SB/SE) and Large Business and International (LB&I) Divisions to develop and report on analyses that could provide stakeholders with more confidence in the methods used to estimate undetected noncompliance, and information on its root causes.

Comment:
IRS agrees with this recommendation. RAAS is conducting both descriptive and root cause analyses of DCE estimates in conjunction with SB/SE and LB&I that will provide stakeholders with more confidence in the methods used to estimate undetected noncompliance and identify root causes.

Proposed corrective action deliverable date March 15, 2025.

Recommendation 2:
The Commissioner of Internal Revenue should fully define the roles and responsibilities for implementing the National Research Program (NRP) Artificial Intelligence (AI)-informed sample selection process.

Comment:
IRS agrees with this recommendation. RAAS revised existing documentation further defining roles and responsibilities which was shared with GAO on April 2nd. IRS is waiting to see if this recommendation is removed from the final report.

Recommendation 3:
The Commissioner of Internal Revenue should complete documentation on key components of the NRP AI models, such as which model options were considered and chosen for implementation, and how IRS determined risk levels and time frames for risk level updates.

Comment:
IRS agrees with this recommendation. RAAS will update the existing documentation that describes the options that were considered for the NRP AI models to specify which model was implemented and discuss time frames for risk level updates.

Proposed corrective action deliverable date July 31, 2024.

GAO-24-106449

Recommendation 4:
The Commissioner of Internal Revenue should complete documentation for the NRP AI models on the technical specifications needed to run the model, such as how the data are to be split or divided.

Comment:
IRS agrees with this recommendation. RAAS will revise existing technical documentation to expand technical specifications outlined in the document to include additional information such as how data are to be split or divided.

Proposed corrective action deliverable date July 31, 2024.

Recommendation 5:
The Commissioner of Internal Revenue should complete documentation for the NRP AI models covering the data sources needed to run the model, such as the list of input variables.

Comment:
IRS agrees with this recommendation. RAAS provided additional documentation to GAO on March 26, 2024, and April 2, 2024. GAO is considering removing this recommendation from the final report.

Recommendation 6:
The Commissioner of Internal Revenue should ensure it has complete documentation on how to update the NRP sample selection process in response to changes in the operating environment, including the performance of its AI models and risk level analysis.

Comment:
IRS agrees with this recommendation. RAAS will update existing documentation to discuss how to update the NRP sample selection process in response to changes in the operating environment.

Proposed corrective action deliverable date February 28, 2025.

Recommendation 7:
The Commissioner of Internal Revenue should complete documentation on processes for assessing the results of AI model updates, including validating the models.

Comment:
IRS agrees with the recommendation. RAAS will update existing documentation to complete the documentation including a discussion of assessing the results of the model and validation.

Proposed corrective action deliverable date February 28, 2025.

2 of 3

GAO-24-106449

Recommendation 8:
The Commissioner of Internal Revenue should develop and document a plan for evaluating the redesigned sample selection process, including factors it will consider in determining whether or how to continue using it.

Comment:
IRS agrees with the recommendation. RAAS will develop and document a plan for evaluating the redesigned sample selection process.

Proposed corrective action deliverable date February 28, 2025.

3 of 3

Appendix V: GAO Contact and Staff Acknowledgments

GAO Contact

James R. McTigue, Jr., 202-512-6806 or McTigueJ@gao.gov

Staff Acknowledgments

In addition to the contact named above, the following individuals made contributions to this report: Neil Pinney (Assistant Director), Robyn Trotter (Analyst in Charge), Chinmay Amin, Arpita Chattopadhyay, Michele Fejfar, Rob Gebhart, Jackson Gode, Krista Loose, Melanie Magnotto Win, Dan Mahoney, Carl Nadler, Andrew Olsen, Andrew J. Stephens, Alicia White, Khristi Wilkins, Sarah Steele Wilson, and Mercedes Wilson-Barthes.

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