TECHNICAL

NOTES & MANUALS

The Revenue Administration Gap Analysis Program

An Analytical Framework for

Personal Income Tax Gap Estimation

Fiscal Affairs Department

The Revenue Administration Gap Analysis Program: An Analytical Framework for Personal Income Tax Gap Estimation

Prepared by Mick Thackray (Fiscal Affairs Department), Sarah Jennings (HM Revenue & Customs), and Martin B. Knudsen (Danish Tax Agency)[1]

Authorized for distribution by Victoria Perry and Ruud De Mooij[2]

CONTENTS

I. HOW DO COUNTRIES MEASURE NONCOMPLIANCE IN PERSONAL INCOME TAX?

It is generally not easy to measure revenue not collected due to noncompliance, but an increasing number of countries are doing so. By their very nature, noncompliant behaviors are unlikely to be declared by taxpayers, and may well be deliberately concealed. Consequently, they are not easy to quantify through direct observation or survey. Even still, the fiscal impacts of noncompliance are of critical interest, not just to tax administrations, but also to finance ministries and other stakeholders, and a growing number of countries now regularly produce and publish estimated revenue losses due to noncompliance. Good tax gap analysis enables the detection of changes in taxpayer behavior by consistent estimates over time.

According to the International Survey on Revenue Administration,[1] 20 countries measure their personal income tax (PIT) gap,[2] but methods, data sources, and robustness of results vary. PIT gap estimates should include in their scope all taxpayers who have PIT liabilities, whether through earned sources of income via employment or self-employment, or unearned income, such as capital income. Estimates can be made whether the tax is collected directly from the individual or by withholding by third parties (for example, employers' deduction of payroll taxes from employees' wages).

A. What Estimation Approaches Are Applied?

Of the various approaches to PIT gap estimation, random sampling of taxpayers for audit provides the most robust gap estimation method (Figure 1). In general, these estimates extrapolate from non-compliance identified in revenue agencies' audits of taxpayers to arrive at a population value. This bottom-up approach is more often used because a top-down approach, using macro-economic data to model the potential tax and calculate the gap between that and actual collections, is not generally appropriate for PIT. Risk-based audits might be used (in principle, with suitable corrections for measurement bias). However, a properly conducted random audit program (RAP) is preferred from a statistical standpoint and has other benefits that outweigh its additional costs. This is discussed in more detail below.

B. Why Measure the PIT Gap?

Tax gap analysis provides tax administrations and their stakeholders with a measure of the amount of tax revenues lost or foregone through both non-compliance and policy decisions. While a modern tax system is predicated on voluntary compliance, there are often few other tools available to a revenue administration to measure and monitor taxpayer compliance holistically. Tax gaps estimated using the IMF's RA-GAP approach, provide a comprehensive analysis of overall

---

[1]   International Survey on Revenue Administration (ISORA) is a multi-organization international survey that collects national-level information and data on tax administration. See International Monetary Fund (2019) and Organisation for Economic Co-operation and Development (2019).

[2]   In addition, 35 OECD countries reported conducting random audits in 2017 (not limited to PIT), of which 14 used the resulting data for tax gap estimation. See Table A.146 Tax gap and Table A.170 Random audits in Organisation for Economic Co-operation and Development (2019).

FIGURE 1. Estimating the PIT GAP from Random Audits

Total PIT gap	=	Average audit adjustment	×	Taxpayer population
	=	Total audit adjustments	×	(Taxpayer population / Sample population)
Where,		Average audit adjustment	=	PIT liability − PIT paid
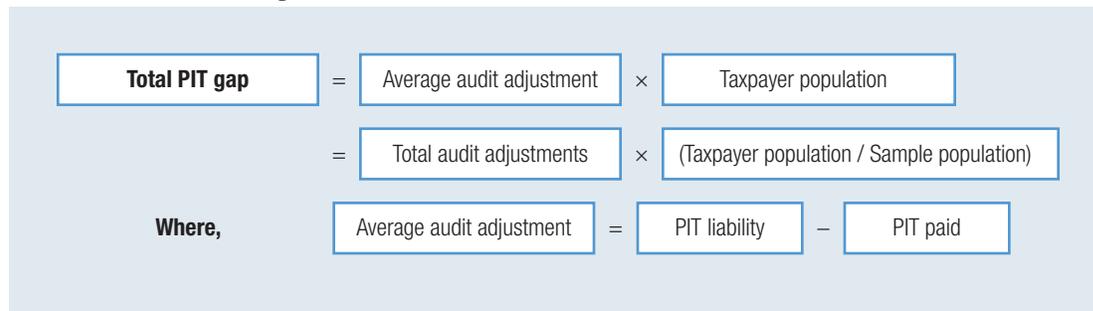
revenue losses that can be used as strategic context for decision making on resource allocation and individual administrative or policy measures. They also contribute to transparency in public administration as a strategic performance indicator for tax administration.

Using a bottom-up approach to PIT gap estimation can produce very rich, detailed results. Bottom-up approaches rely on observations of actual compliance behavior by individual taxpayers. Although such approaches are more demanding of time and resources than top down approaches, the data produced can provide a detailed analysis of compliance risks and components of the compliance gap for use by RAs in compliance risk management. This is particularly the case for RAPs, where a well-designed program can provide a foundation for an evidence-led risk management strategy (see Table 1).

C. The Additional Benefits of Conducting Random Audits

In addition to the measurement benefits (see below), such programs can yield substantial additional revenue. Random selection of some taxpayers for audit is a recommended part of an audit strategy and can increase the deterrence effect where the audit plan itself is published.[3]

Selecting cases randomly means all parts of the taxpayer population are tested for risk, which improves risk profiling and compliance results. There is a danger where risk-profiling is built on data from risk-based cases only that the predictive power is only as good as the original understanding of risk—i.e., cases continue to be selected from the same, limited pool of taxpayers who have characteristics similar to those already identified as risky. Having an additional source of data from random audits can help counteract this and improve risk-profiling.[4] The random audit program also provides results that can be used as a counterfactual when evaluating how effective risk assessment is in selecting cases for risk-based audit.

---

[3] See Biber (2010) for the IMF's Technical Note and Manual Taxpayer Audit— Development of Effective Plans, which notes the benefit of publishing plans for risk-based audits.

[4] See Dellaportas et al. (2019) for a methodological approach to estimate the optimal blend of random and risk-based audits to achieve revenue benefits through refined risk-profiling.

Table 1. Costs and Benefits of Random Audit Programs

Benefits	Costs
• Statistically robust methodology	• Requires compliance staff resources
• Relatively simple analysis	• Time needed for first estimate (1–2 years lag)
• Quantifiable error	• Requires adjustment for unregistered taxpayers and undetected risks
• Evidence-based risk and behavioral insights	• Opportunity costs of random audits replacing risk-based audits
• Improved risk profiles and better targeting of risk-based audits	
• Additional deterrence boosts voluntary compliance	

D. Structure of This Note

The following sections set out the theoretical concepts for PIT gap estimation, the different measurement approaches available, and their implications for scope and presentation of statistics. The final section focuses on the practical steps for measuring the PIT gap by establishing a random audit program to collect data, and how to scale findings from the sample to the population. Further technical information is provided in the appendices.

II. WHAT IS A PIT GAP?

As defined in the IMF Revenue Administration – Gap Analysis Program (RA-GAP),[5] the tax gap is the difference between potential tax revenue from the tax base and actual revenue. The tax gap can be decomposed into two main components: losses from non-compliance (the *compliance gap*), and the impact of policy decisions (the *policy gap*).[6] Defining the gap in this way allows for a comparison of the sizes of the compliance gap and policy gap and their respective contribution to the total gap. This allows tax administrators and policy makers to assess potential avenues of actions for improving revenue performance by addressing either component of the gap. This note focuses mainly on the compliance gap.

FIGURE 2. Illustration of the Components of the Tax Gap

The recommended RA-GAP approach for estimating the gap is either 'top-down' or 'bottom-up' depending on the tax. Top-down gap estimates use independent statistical data, typically from national accounts, to estimate total potential tax, against which actual tax collections can be compared. A top-down approach can allow for estimation of both the compliance and policy gap, and results in estimates that theoretically include all revenue losses.[7] In contrast, bottom-up gap estimates are derived from on observations of individual taxpayers' noncompliance.

---

[5]   See Box 1.

[6]   Hutton (2017).

[7]   See Hutton (2017) and Thackray and Alexova (2017) for a comprehensive discussion of the theoretical framework and requirements for a top-down approach.

A bottom-up approach is usually more appropriate for estimating PIT compliance gaps than a top-down approach. For PIT, the administrative data used for compiling the actual tax revenue is very often also the primary source for establishing personal income or its distribution in national statistics.[8] Without an independent source of data, a bottom-up approach is typically required to estimate the PIT compliance gap.[9]

<div style="background-color:#faf0c8;padding:1em;">

BOX 1. IMF RA-GAP Program

The Revenue Administration Gap Analysis Program (RA-GAP) is a capacity development service for IMF member countries, conducted by the IMF Fiscal Affairs Department's Revenue Administration Divisions (FADR1 and FADR2). It provides revenue administrations with comprehensive and detailed estimates of the gap between current and potential collections, as well as a review of current operational performance in a number of key functions.

The goal of RA-GAP is to estimate the tax gap and identify some of the underlying causes of the gap. While the tax gap is a crucial key performance indicator (KPI) for a revenue administration's overall effectiveness in collecting tax revenues, it is as important to be able to identify what is contributing to the gap.

Previous technical notes published by the IMF have documented the RA-GAP analytical approach to estimating tax gaps for VAT (Hutton 2017), CIT (Ueda 2018), and excise taxes (Thackray and Alexova 2017).

</div>

A. The PIT Policy Gap

There is no international consensus on how to define a standard tax base for PIT. The varying level of progressivity in PIT regimes and different types of rule-based deductions make it difficult to agree on a baseline (i.e., a single, standard rate or theoretically 'ideal' regime) for establishing tax expenditures for PIT.[10] The non-taxable part of the policy gap includes incomes that are inherently difficult to identify and quantify, so are not included within most potential PIT bases. An example would be gifts from family or friends. There is no international agreement on the basic filing unit – the individual or the family; and this makes establishing the baseline more difficult.

Given the variation in definitions of both the standard rate and the reference policy structure, RA-GAP does not provide a definition of the PIT policy gap. It is not possible to do so in a way that is both applicable to virtually any country for comparison purposes, while remaining

---

[8] Although this is not universally the case, and the balancing of the three GDP measures means that total income reported in national accounts is calibrated to independent sources, this is a major stumbling block in practice in most countries. For an exception, see Galucci et al (2020) for a top down methodology the Italian Revenue Agency is developing as an indicator of the PIT gap. See also Rubin (2011) for a discussion regarding the practicalities of a top-down approach for estimating tax gaps for direct taxes.

[9] Further discussion of the suitability of a top-down methodology for estimating the PIT gap is provided in Box 2 of this paper.

[10] The lack of an agreed baseline for PIT is also an issue in establishing consistent measurements of tax expenditures.

meaningful to individual countries. These definitional difficulties mean that even where an independent source of data on the PIT economic base exists, PIT is less well suited to top-down gap analysis.

These definition issues for a PIT are in contrast to VAT, and to a lesser extent CIT, where there is broader consensus as to the reference policy structure. The latter tax liabilities are also generally less contingent on the individual taxpayer's personal circumstances than PIT. For these reasons, this note does not address the measurement of the policy gap for PIT but focuses exclusively on measuring the compliance gap for PIT using a bottom-up methodology. In terms of Figure 2 (above), the compliance gap is still defined as the difference between actual and potential revenues given the current PIT policy.

III. HOW CAN THE PIT GAP BE MEASURED?

A. An Overview of Approaches

The methods discussed below can be applied to both withholding and self-assessment PIT regimes.

Estimating from Random Audits

Random sampling is a means of selecting taxpayers for audit in which every member of the population of interest has a known, non-zero probability of selection. Following a random selection method means the sample is theoretically representative of the full population. The sample is subject to a comprehensive audit[11] and the auditors record the value of adjustments made (the difference between the taxpayer's and auditor's assessment of liabilities).[12] The adjustment is then scaled by the ratio of sample to population to quantify the gross tax gap, adding in any adjustments to account for non-detection, non-payment, and the hidden economy (see below). There are different types of random sample designs that can yield representative results and several issues to consider in order to accurately scale results to the population, described below and in Appendix A.

Estimation from random audits generally provides the most robust results and is the recommended approach of RA-GAP for PIT gap analysis. Reliable estimates require a well designed and implemented program, but also depend on the nature and distribution of compliance risks. In addition to the benefits discussed above, a random sampling design allows for quantification of the precision of estimates in terms of sampling error.[13] However, it may not be a suitable approach for very small populations as a large proportion would have to be audited for reasonably precise results. While there are measures that can help boost the revenue return on investment of a random audit program, discussed in section IV, it remains a resource intensive approach (see Table 1).

Estimating from Risk-Based Audits

Risk-based audits can be used to estimate the PIT compliance gap using methods that attempt to correct for selection bias, but come at a cost of either precision or certainty. To avoid the additional costs of random audit programs, revenue administrations (RAs) may try to estimate the PIT gap from the results of their existing, risk-based audit program, but this is problematic. Risk-based audits target taxpayers with the highest assessed or assumed risk and are therefore inherently subject to selection bias. If the results from risk-based audits were scaled

---

[11] For the purposes of this note a tax audit is defined as an examination to determine whether a taxpayer has correctly reported and assessed their tax obligations (Biber 2010). In the case of random audits, an audit is defined as a 'comprehensive audit', meaning the examiner checks all areas of potential risk as opposed to focusing on specific risks or limited items on the taxpayer's filed return. Examining only some risks/items would lead to an underestimate of the PIT gap and limit the insight gained from the RAP.

[12] In a top down approach, the total tax base is estimated from aggregate statistical data. When using audit results, this calculation is made in the course of each audit: and the potential tax base under the current policy framework is estimated by the auditor, as they establish what the taxable income should be.

[13] Other types of error, for example, from assumptions used in models, are not quantifiable.

directly to the total population without taking this into account, they would likely over-estimate the tax gap.[14] In order to extrapolate from a risk-based sample to the population, this selection bias must be corrected.

BOX 2. Estimating the PIT Gap in Low Income Countries

The costs of a RAP are better justified for low income countries by the improvements it can bring to compliance risk management and risk-based audits than simply as a way to estimate the PIT gap. The apparent costs of an RAP can be off-putting, particularly in low income countries; and difficult to justify solely as a means of estimating the compliance gap. But the main benefit of RAPs is perhaps not in the measurement of the overall gap itself but that they can be used in risk profiling, to monitor and improve risk-based selection for audits. As such, they are a recommended component of evidence-based audit programs. Where auditing capacity is insufficient for a separate RAP, random sampling can be integrated into case selection for the existing audit program. In this approach, some of the cases chosen for audit are selected randomly from the total taxpayer population, and their results identified and used to estimate the total compliance gap.

For low income countries, top-down approaches to PIT gap estimation, based on national accounts data may appear more attractive. Typically, the informal economy, and associated compliance gap, is larger in lower income countries.[1] As well, national accounts in low income countries are less likely to rely on data from tax records.[2] As a consequence, compliance gap estimates may be more readily obtained from a comparison of aggregate incomes and PIT collections. However, the accuracy of these estimates will depend critically on how well the national accounts capture unreported income and its distribution among taxpayers, as well as the detail of other factors potentially affecting individual taxpayers' liability,[3] and this needs to be reviewed critically.

Similarly, aggregate incomes from household income surveys may offer a top-down approach to estimating the PIT base for gap estimation. Where available, such surveys generally offer an analysis of aggregate households' income and its distribution that might be used to model the PIT base. However, their estimated totals for all households are very often distorted by non-response (particularly for households at the high and very low end of the income distribution) and under-reporting by respondents.[4] They therefore need to be calibrated to national accounts or other aggregates (e.g. tax collections) to adjust for such biases, exposing them to much the same limitations in gap estimation.[5]

[1] According to the experience RA-GAP has gained through capacity development missions estimating tax gaps for member countries.
[2] See Rivas and Crowley (2018).
[3] See Quiros-Romero and others (2021).
[4] Hurst and others (2010), who find that self-employment income is underreported on household surveys.
[5] This technique is described for Latin American countries by Pecho Trigueros and others (2012), who note processing the data may be "laborious."

---

[14] Unless the risk-based sample selection is not an improvement over random selection, or under-detection of non-compliance in risk-based audits is severe.

In the absence of existing RAP-based estimates, there may well be value in an initial assessment of the PIT compliance gap using the top-down approach. Where the expected PIT gap is relatively high, a broad estimate of its scale still provides a useful strategic risk indicator (and can inform the design of a RAP). The levels of uncertainty associated with national accounts aggregates are less important for larger gaps (i.e. it is easier to tolerate a margin of error of plus or minus five percentage points for a compliance gap of 30 percent of liabilities than for a 5 percent gap).

Methods to model the tax gap based on risk-based audits include:

- **Post-stratification:** taxpayers (both audited and non-audited) are divided into strata using variables that are relevant in the selection of cases for audit, the tax gap for each stratum is estimated by scaling stratum audit results to the total population size for each stratum.[15]

- **Heckman two-stage procedure:** The model consists of two steps. The first estimates the probability that a case will be included in the sample (i.e., selected for an audit). The second step models the outcome of an audit (i.e., the determinants of yield) using both explanatory variables and a regressor that controls for the selection bias. This regressor is calculated based on parameters estimated in the first step.[16]

- **Extreme value method:** The method is based on the assumption that the majority of non-compliance is identified by existing audits of higher-risk taxpayers and extrapolates from these to infer the value of residual non-compliance in lower-risk, unaudited taxpayers. It can be applied where the distribution of the data follows a particular power law—a Pareto distribution—whereby a small number of 'extreme' cases account for the majority of the total value of revenue found in risk-based audits.[17]

Each approach has its drawbacks, requiring either high levels of data on taxpayer characteristics and selection criteria for audits, or strong assumptions to be upheld. The scale of error in estimates is largely unquantifiable, but results may be comparable between countries where the procedures are applied consistently. Apart from the Heckman two-stage approach, an increase in hit-rate will be interpreted as an increase in non-compliance in the population. Trend analysis will therefore be misleading if the improved hit-rate actually reflects improved capability of the RA.

Indirect Estimation Approaches

Partial estimates of the PIT gap, generally related to under-reporting of income by the self-employed, can be estimated through indirect modelling techniques. For example, the income-expenditure approach models under-reporting of income through observation of the relationship between consumption (generally food expenditure) and reported incomes.[18] A case study of

---

[15]   The Italian Revenue Agency has employed this technique for estimating their CIT gap, see European Commission (2018). The method has also been adapted, for example, by using machine learning techniques, such as cluster analysis, to group taxpayers into groups with similar characteristics. See Canada Revenue Agency (2019).

[16]   See Heckman (1979). See European Commission (2018) for an elaboration on both the Heckman method and post-stratification approach for tax gap estimation, including key strengths and weaknesses.

[17]   See Bloomquist et al. (2014). For application of the method, see Internal Revenue Service Research (2019), Canada Revenue Agency (2019), HM Revenue and Customs (2020) and Australian Tax Office (2020).

[18]   See Pissarides and Weber (1989), Bazzoli et al. (2020), Engström and Hagen (2015), and Gonzalez Cabral et al. (2019) for examples of its application and discussion of approaches. See also Hurst et al. (2010), who find that self-employment income is underreported on household surveys.

this approach by the Italian Ministry of Economy and Finance is contained in Box 3. Whilst the approach is conceptually attractive, the process of preparing and manipulating the data can be very laborious. Estimating the gap will still require a number of assumptions to be made, so the resulting estimates are subject to an unknown degree of error and will not be comprehensive PIT gap estimates, though they may still be useful to improve RAs' understanding of components of the gap, for example tax not paid by non-filers.) Given the high analytical requirements, indirect estimation approaches may be best undertaken by external experts, such as academics, commissioned by the RA.

BOX 3. Case Study: Estimating the Italian PIT Gap Using the Income-Expenditure Method

In 2019, the Italian Ministry of Economy and Finance published bottom-up estimates of the PIT gap to provide a more detailed description of self-employed income under-reporting and related distributional analyses.[1] The method adopted is an extension of the Pissarides-Weber approach that allows for calculation of detailed breakdowns of income evaded by the self-employed by different categories of households (e.g., single vs married, location of residence, presence of children, etc.).[2] The richness of the dataset is a key feature of the Italian model: household consumption data from the Household Budget Survey for the year 2013 were individually matched, through an anonymized personal identification number, with income data from income tax declarations for the years 2009-2016. Data on property and financial wealth from administrative sources were also used to reconstruct true income more precisely.

The observed share of unreported income by self-employed taxpayers was more than 40 percent of their income on average. Differences in evasion rates were observed for particular household characteristics. The results for Italy suggested that, for instance, the share of undeclared income is less for in-couple than single head households. The findings also indicated that self-employed households with a college educated head tended to evade more than similar wage and salary households.

The calculation of the potential PIT base and PIT tax gap makes use of EUROMOD, the EU-wide tax-benefit microsimulation model.[3] The policy rules used were for 2018, with income values from 2013 tax data uprated to 2018. The resulting gap for self-employment for the year 2018 was about €34 billion, which is the sum of about €26 billion of PIT gap and €8 billion of the social contributions paid by the self-employed. These estimates are in line with top-down PIT estimates that are available for the same year, about 67 percent of gross tax liabilities. More than 80 percent of missing tax revenue is attributable to the taxpayers at the top of the income distribution. In considering the distribution of income if there were no under-reporting, reported income inequality in Italy would be higher. However, the redistributive impact of the tax-benefit system would be greater.

[1] The bottom-up estimates are produced by the Italian Ministry of Economy and Finance and complement the existing top-down estimates made by the Italian Revenue Agency, see Galluci et al. (2020). Top-down estimates also contain some information from risk-based tax audits, see Ministero dell'Economia e delle Finanze (2019).
[2] See Bazzoli and others (2020).
[3] See Ceriani and others (2020).

Other indirect estimation techniques, such as a Multiple Indicators Multiple Causes (MIMIC) estimation, are not recommended for revenue administrations. MIMIC modelling produces an estimate of the size of the shadow economy based on a number of observable indicators that are assumed to provide an indirect measure of hidden economic activity.[19] Properly executed and reported, such studies can provide useful social indicators; but, although shadow economy activity can contribute towards the PIT and other tax gaps, they are not synonymous. MIMIC approaches also necessarily rest on strong assumptions that are not always justified. This estimation approach has been subject to considerable criticism as a means of measuring compliance losses and is not recommended by RA-GAP.[20]

There are other econometric and survey approaches, but these also tend to be studies of the shadow economy and not of tax non-compliance specifically.[21] Given that perpetrators of tax non-compliance often take deliberate steps to conceal their behavior, it can be hard to observe non-compliance directly. Consequently, econometric and survey approaches are very often conceptual and/or heavily assumption driven. Their value depends on how closely the available data measures the concept in question; how reasonable the treatment of the data and assumptions used are; and how accurately results are reported with respect to limitations and uncertainty. There is always a risk in the generation of estimates of the shadow economy that the results may be misreported by the mass media or misused to suit political agendas. Extreme caution is therefore advised in adopting any of these approaches for tax gap estimation.

B. What Is in the Scope of the PIT Gap?

The scope of the estimated gap will depend on definitions used, sample design, and methodological approach taken. The compliance gap defined by RA-GAP is the difference between the current potential tax under existing policies and the actual tax collected. However, the final scope of a PIT gap estimate may be partial, reflecting: 1) practical limits to the data that can be collected from a random audit program or is otherwise available to the analysts; and 2) what is cost effective to include in scope. It is important in reporting PIT gap estimates to clarify what is, and is not, covered by the gap estimate.

The tax gap should be for a specified time period. Typically, the time period will be the fiscal year and should align with the period typically referred to by the RA when reporting its performance. Although seemingly obvious, this is important to clarify as when using data from audits, auditors may investigate non-compliance they find across multiple years, which could result in the gap being overestimated if the data cannot be disaggregated. Some RAs publish periodic estimates covering multiple years, particularly if running a random audit program, where the sample of audits can be collected over several years to mitigate the costs of large-scale programs.

Estimates will not be comprehensive estimates of the PIT compliance gap without adjustments for unregistered or non-filing taxpayers. Audits, whether random or risk-based, are generally of registered taxpayers or filed returns. Where this is so, the PIT liabilities arising from unregistered or non-filing taxpayers will not be directly covered. Additional methods may need to

---

[19]   See Schneider and others (2010).

[20]   See, for example, Breusch (2016) and International Monetary Fund (2015).

[21]   See, for example, the range of approaches discussed in Pecho and others (2012).

be developed to take account of PIT due from the non-filing population, such as research surveys into hidden economy activities or data-matching to third party information on unreported sources of income.[22]

Results of audits should be adjusted for non-detection bias, ideally taking into account differences in the capacity of auditors and/or the scope of audits. Multipliers to apply to observed results can be developed to allow for non-detected risk and reach a theoretically comprehensive PIT compliance gap. However, this is methodologically challenging—the results can be highly uncertain and will add an unquantifiable error to the gap. In the absence of a multiplier, the gap can be described in reported findings as a lower-bound estimate or the detectable PIT gap. Appendix B discusses different approaches to deriving non-detection multipliers.

There are specific types of low frequency but high value risk that a random audit program is unlikely to capture reliably. The reason for this could be that they are too infrequent to be caught in a reasonably sized sample, or that, without prior information, routine audit techniques will not uncover more complex or deliberately disguised frauds. A supplementary estimate or adjustment can be developed to cover these components of the PIT gap.[23]

Using the RA-GAP approach, the compliance gap is estimated net of revenue from voluntary adjustments and compliance interventions, i.e., the amount of tax that will never be recovered. RAs may also calculate the gross gap, including PIT not paid on time, based on returns and payments received by the due date, and then deduct late payments received to calculate the net gap.

PIT compliance gap estimates from audits are wider in scope than just intentional fraud and evasion. Estimates will also include under-payments due to unintended taxpayer error, and the offsetting effects of over-payments (though see Box 4). They may include amounts attributable to avoidance or aggressive tax planning that the revenue administration seeks to challenge. However, in practice, the larger and more complex avoidance schemes will likely be too infrequent (or difficult to identify) to be captured reliably in random audit programs.

C. What Can We Say about Precision?

All tax gap estimates are subject to error, so the PIT gap results should be interpreted as a broad estimate. However, so long as the estimation approach and data used are consistent over time, the scale and direction (i.e., over- or under-estimation biases) of error should be roughly constant. Consequently, though there is uncertainty in the level of any single estimate, good tax gap analysis provides robust estimates of the trend over time. This enables the detection of changes in taxpayer behavior by allowing meaningful comparisons to be made between estimates over time.

---

[22] See Organisation for Economic Co-operation and Development (2002) for a discussion of methods and Doyle and others (2017) for an example research survey into hidden economy activities.

[23] See Canada Revenue Agency (2018) for an approach to estimating the tax gap associated with hidden offshore investments.

BOX 4. Are Unintentional Taxpayer Errors Unbiased?

Intuitively, genuine errors might be expected to be as likely in one direction (resulting in too little tax being paid) as the other (too much tax being paid). So the net impact of unintentional error by taxpayers could therefore be expected to be unbiased—that is, made to a similar extent and value whether over or under-declared. However, in practice it is generally found that unintentional errors are more likely to increase the observed compliance gap than otherwise.

This could be because taxpayers take more care not to overpay taxes than to underpay but there may be another factor involved. The finding of an unintentional error by an auditor is likely to generate less paperwork (and make for a quicker audit) than a finding of deliberate evasion. This could create a perverse incentive for auditors to under-report the seriousness of taxpayers' motivation, and further bias the observed impact of unintentional errors upwards.

Mathematical measures of precision generally only capture sampling error, whereas tax gap estimates will likely be more affected by non-sampling error. Non-sampling error includes bias in the final sample and errors made in analysts' assumptions. This is generally not possible to measure, although sensitivity analysis of assumptions can provide insights into their potential impact on error. It may not be helpful to present mathematical measures of margins of error or use statistical significance test in final reports as these can give a misleading view of precision.[24]

D. How Should the PIT Gap Be Reported?

It is best practice to primarily report the PIT gap as a proportion of the potential PIT. It can either be expressed as a percentage of total liabilities (where the value of the net gap is divided by total liabilities). Focusing on the cash value is less useful, particularly in analysis over time where the cash value may rise or fall in line with liabilities in a changing economy.

It is good practice, where feasible, to provide evidence-based explanations of changes in trends over time. More granular level analysis of specific risks and behaviors should support this, particularly where estimates are based on random audits. With this insight, analysts may also be able to offer some interpretation to assist decision makers in strategy design.

Publishing tax gap estimates promotes transparency and accountability. Gap estimates inform public debate by providing robust quantification of the scale of non-compliance. Publishing estimates, explanations, and associated methodologies also allows for review and feedback by external experts, which can be helpful for improving tax gap estimates in future (see Box 5).

The RA should be prepared to be held to account as to the size of the PIT gap and any changes in it over time. Not all of the compliance gap is a consequence of the RA's performance or within their control. For example, some tax owed is unrecoverable as taxpayers

---

[24]   For example, the Internal Revenue Service does not report these measures of precision for this reason. See Internal Revenue Service (2019).

become insolvent, and the rate of insolvency can change due to external factors such as a recession. However, the tax gap remains a key indicator of the RA's performance and will likely be scrutinized.

<div style="background-color: #faf0cc; padding: 1em;">

BOX 5. Tax Gap Estimates as Published Statistics

Following published statistical governance frameworks, either at the country or international level, and gaining an official status can be a benefit for enhancing trust and credibility of published estimates. This is particularly helpful given the revenue agency will come under scrutiny by the public, media, and elected representatives when releasing estimates.

As part of this, having a clear policy on how and when statistics are revised is beneficial. In the United Kingdom, HM Revenue and Customs publishes its policy on revisions to statistics. This policy applies to tax gap estimates as they have Official Statistics status. This is in line with guidance issued by the UK Statistics Authority, the body which governs the Code of Practice for Official Statistics, and has the objective of enhancing trust and transparency in statistics. See HM Revenue and Customs (2010).

</div>

IV. WHAT ARE THE STEPS FOR ESTIMATING THE PIT GAP FROM RANDOM AUDITS?

FIGURE 3. Summary Process Flow of a Random Audit Program

| Analysts explore population data, scope for stratification, and probable ranges of precision | Agree sample size, balancing precision of estimates against resources | Prepare sample frame (population list), checking it represents population of interest | Segment sample frame by strata and select using a random sampling method | Sample sent to auditor teams for compre-hensive audit | Regular monitoring and quality assurance of audits and data recording | Extract data from the program for estimation (see Figure 3 below) |

A. Sample Size and Design

Different random sample designs can yield valid results so long as every taxpayer in the population of interest has a chance of selection. Where resources are constrained, integrating some random case selection into the general audit program of the RA is an option for generating useable data while minimizing resource costs. Different random sample designs are set out in Appendix A.

The sample size is calculated based on the degree of precision that is acceptable and resources available. This will need to be calculated for any subgroups for which the RA wants detailed estimates, as well as estimates for the whole population. The RA will need to balance the levels of precision of estimates and lower-level insight (into particular risks or subgroups), against audit results (in terms of revenue collected) and resource costs (e.g., staff costs). Analytical teams can set out the margins of error expected at different sample sizes to demonstrate the impact on precision when agreeing resource for the random audit program (see Appendix C for an example). Beginning with a pilot study of a smaller size sample is recommended for understanding the distribution of non-compliance and better estimating the requirements of the full-scale program to produce sufficiently precise results. A pilot study also helps test procedures prior to investing significant resource in the program.

If a sufficient sample size cannot be resourced in a single year, spreading the program over multiple years and pooling the results can be considered. This has the advantage of achieving a higher final sample size at a lower annual cost. The resource to select the samples each year and manage the program may be higher (compared to a larger program run in a single year). This inevitably delays the production of the full estimate by a considerable period, potentially several

years, but interim or provisional findings may be produced in the meantime. Estimates will refer to a broader time-period, which is a disadvantage especially if any significant changes to tax policy, operations, or the economy occur part-way through the period. It will not be possible to disaggregate year-on-year changes with precision if annual sample sizes are low.

Stratifying the sample can yield both analytical and revenue benefits from the program. If subgroup analysis is desirable, for example, by business sector or income groups, segmenting by groups of interest will help assure sufficient cases are selected from each group. In order to achieve this, the taxpayer registry data will likely need to be joined to returns data to gain data on characteristics of taxpayers to use in specifying strata. This can limit the sample frame to those who have previously filed, so can be a source of bias if those who have registered and not yet filed or failed to file are excluded.

If risk indicators exist for the population, it is recommended to stratify based on risk—even in the first random audit program. Not only does this allow testing of the risk profile, and help assure full coverage of identified risks, but it can increase both the precision of the final results and revenue yield from the random audit program. This type of stratification could be as simple as segmenting the population into 'high' and 'low' risk categories and over selecting the sample from 'high'. Without risk indicators, segmenting by turnover or income and over selecting from those in the highest earning segment is often most appropriate. This will tend to increase the precision of estimates by increasing sampling probability in the high risk/high income groups, where greater variance is expected.

B. Selecting the Sample

Sample selection requires that a list of the population is available, which is generally the list of taxpayers in the RA's registry. The list, or sample frame, should be comprehensive of the population of interest and as up to date and accurate as possible. It can be advantageous to initially select a larger sample than required for the final results so there are spare sample units available for replacing deselected cases. This also allows for project expansion should there be some flexibility in resource. Selecting an additional 'top-up' sample after the program starts will not necessarily introduce bias if the population definitions remain unchanged, but can mean more resource is spent on preparing an additional sample. Checking the sample frame against other data sources to confirm it is relevant, for example, checking for any changes to the taxpayer's status, or emigration or death, will help save wasting resource in the program.

In the absence of a population list, a systematic sample can be selected based on every Nth filer, based on a random start number. This is disadvantageous as it creates uncertainty as to the resource requirement for the program and can make selection by strata more difficult. It can introduce bias if the program is stopped before the entire population for the period of interest has filed, or if there is a sequence to the filing data itself.

Selecting the sample quickly and progressing to fieldwork is important if there are legal limitations on the time available to the RA to open an audit. Losing the sample as deadlines have passed can be a significant source of bias, as filing timing may be associated with non-compliance. Selecting the sample quickly also helps to avoid losing cases that are selected for risk-based audit, which would likely bias the results downwards (although this is less of an issue where there is a very large population and low risk-based audit rate). The findings from these specific risk-based

audits can also be brought into the analysis as a substitute for the random audit to avoid this issue. Bearing in mind the timing of filing in plans for the program, and allowing some resource for audits into late filers, will also help reduce bias in the sample.

C. Data Requirements and Recording

Keeping a record of any cases not taken up for audit is information that can be used in estimation. Recording the number of cases that were deselected from the sample as they were found not to be within the population of interest can be used to determine the relevant sample ratio (see formula below) and adjust the population size to more accurately estimate the gap. Recording cases already subject to a risk-based audit, or which should not be taken up for some other reason, can help identify cases where values can be substituted or imputed to avoid introducing bias in the sample.

There is a strong human influence on data quality in audit data. Unlike risk-based audits, the objective of random audits is not hit-rate and revenue, but thoroughness of investigation and high-quality data reporting. Auditors and management will need to have clear guidance and understanding of the objectives of the random audit program to help secure high-quality data. Regular monitoring and quality assurance reviews can also facilitate higher standards of data recording.

Guidance covering standardized reporting of all audit results and other important characteristics will be beneficial to overall data quality and insight. Data captured needs to include positive and negative reassessments recorded separately and a disaggregation of the non-compliance of the period of interest from any other time periods investigated. Other dimensions to consider to achieve greater insight include:

- Codes for specific non-compliance types and behaviors.
- Whether the taxpayer is represented by an agent or accountant.
- The business sector.
- The specific types of income or line items where non-compliance was identified.

It is good practice to record the type of behavior that led to the reassessment. A standardized reporting scale to distinguish between intentional (e.g., evasion) and unintentional (e.g., error) non-compliance could be used to report results. Where possible, this should align to legal definitions, for example, if the RA has a penalties regime based on taxpayer behavior or intention. If resources allow, the RA could consider assessing behaviors or types of non-compliance by post-audit data coding by analysts or specialist teams as this may offer greater reliability of results. Insight into behaviors driving tax non-compliance can be highly beneficial to the development of the RA's organizational and compliance strategy.

Additional Data Requirements

Enforced payments from compliance activity and debt data will be required from the RA's records to estimate the net PIT gap. Unlike estimation for indirect taxes, payments and debt data can generally be taken at an aggregate rather than taxpayer level. Debt is generally added to the

tax gap estimate (as 'non-payment' in the formula below) where it is considered unrecoverable. It is necessary to disaggregate data to the level of the target population, any subgroups, and time period under consideration to align with tax gap analysis.

D. Scaling Results to the Total Taxpayer Population

FIGURE 4. Summary Tax Gap Calculation, Estimating from Random Audits
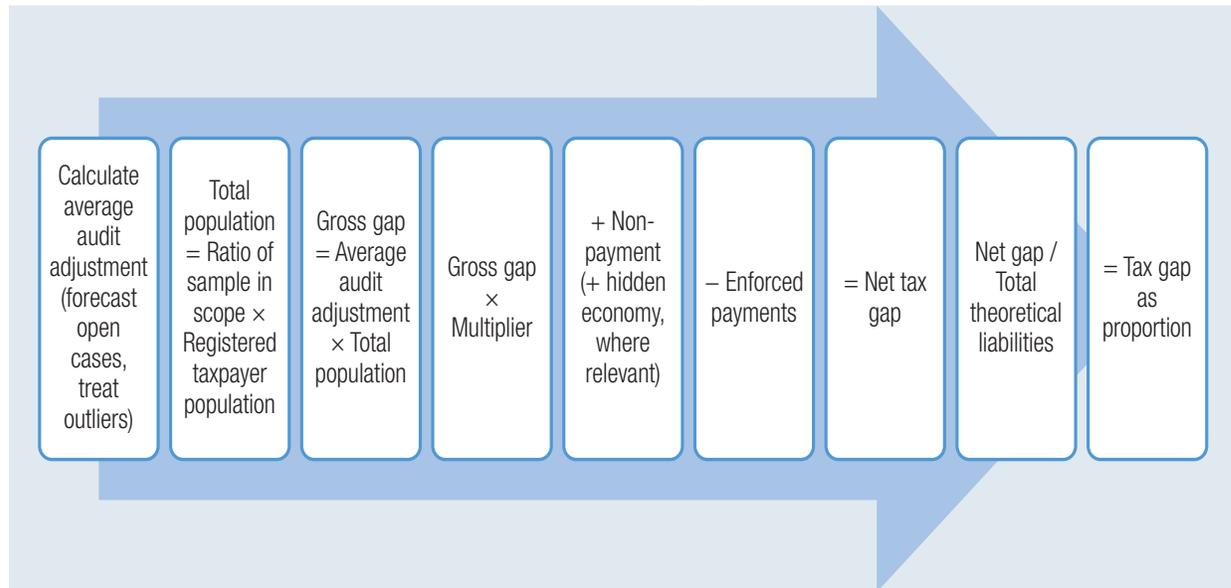
The PIT gap derived from a Random Audit Program for a given period can be expressed as:

$$Net\ tax\ gap = \bar{a} \times \bar{m} \times N_t + d + h - e$$

where,

Net tax gap is defined as gross non-compliance after enforced payments
$\bar{a}$ = average audit adjustment
$\bar{m}$ = average (or aggregate) non-detection multiplier (if used)
N_t = number of filing taxpayers
d = non-payment at time of calculation
h = estimate for hidden economy (non-filers), if available
e = forecast enforced payments (if used)

Notes:

1. Additional adjustments may also be appropriate for gaps in the RAP coverage, for example low-risk taxpayer segments omitted from the sample, or complex frauds.
2. This is a simplified form of the model, assuming a simple (unstratified) random sample. For stratified samples, the formula is more readily expressed as:

$$net\ tax\ gap = \sum_s \left(\frac{N_s}{n_s} \times \sum_i (a_i \times m_i) \right) + d + h - e$$

where,

N_S = number of filing taxpayers in stratum S
n_S = number of audits sampled in stratum S
a_i = audit adjustment for taxpayer i
m_i = multiplier for taxpayer i

Notes:

1. Additional adjustments may also be appropriate (see below).

Adjustments

An adjustment to the total population number can be made based on the ratio of cases that were deselected from the sample during the random audit program. Cases may be deselected because they could not be traced or were found to be otherwise no longer within the population of interest. Alternatively, and where the sample frame does not provide a complete set of the population of interest (for example where systematic sampling is being used), it may be possible to estimate population sizes from other sources, such as population surveys or censuses. Otherwise, a scaling factor may be based on tax revenues by comparing the revenue reported by the sample to total revenue reported by the filing population to the RA.

An allowance may need to be made for audits still open at the time of estimation. A balance needs to be struck between the timeliness of reporting of the PIT gap and the time required to resolve complex or disputed cases. The final audit adjustment value is not only uncertain (in terms of closing for any revenue if the entire sum is in dispute), but likely to be both larger and less easy to calculate than more routine cases. Consequently, such open cases could have a disproportionately high impact on the final estimate.

The value of open audits should generally be forecast to avoid downward bias in gap estimates. If very few cases remain open, with low expected adjustments, these might be ignored, but if significant amounts are likely involved, it is better to forecast results. This should generally be based on the auditor's latest estimate of the likely adjustment. Alternatively, statistical imputation methods can be used to substitute missing values.[25] Initial gap estimates can be re-estimated based on the final data where necessary. It's good practice to have a revisions policy, especially for published estimates (see Box 5).

Outlier cases (those cases with unusually large audit adjustments) can distort gap estimates if they are not corrected. Outliers can be identified through descriptive statistics to explore distribution. There are different ways of treating outliers—they can be retained unadjusted, deleted, capped, or imputed. Consistency in treatment in estimates over time is most important to preserve the value of the tax gap for trend analysis.

---

[25] There is a range of options from basic replacement, such as substitution of mean values, to more sophisticated predictive or simulation techniques, or machine learning, for example 'nearest neighbor' methods. See van Buuren (2018).

V. FINAL REMARKS

The use of a random audit program as described above is the RA-GAP preferred approach to estimating the PIT compliance gap for most countries. Although such an exercise does require a significant investment, both of time and of staff resources, it provides not only a robust strategic performance indicator, but the detailed empirical evidence required to test existing risk profiles, and improve case selection and compliance risk management generally. The benefits of these operational improvements can be expected to more than offset direct and opportunity costs of a RAP.

Other approaches to estimating PIT gaps are sometimes used, but these may not always be appropriate. Top-down approaches, though perhaps initially attractive as holistic estimates, are not usually appropriate in the absence of independent data sources on individual income. Other bottom-up methodologies are quite often tried, typically drawing on data from risk-based audit selections. However, these methodologies require advanced analytical capability and very detailed, high quality administrative data, which—even where they are available—rarely completely negate the need for strong assumptions. 'Big data' and associated advanced data analytics (for example machine learning and very large-scale data matching) are often still quite experimental, and generally better suited to risk identification and profiling than producing unbiased gap estimates, (As well as more often than not being beyond some administrations' current analytical or IT capabilities.)

Good design and execution of a RAP and its analysis can significantly reduce costs and increase benefits. While the accuracy of the results may seem to depend largely on costs—for example the number of audits, their thoroughness (and of course the training of auditors)—good design and rigorous execution can significantly increase the efficiency of the exercise. Stratified samples can be based on existing risk profiles to increase both the average audit yield and the expected accuracy of results for any given sample size. Systemic quality controls and management of the data capture process will improve the reliability and clarity of results. Appropriate multipliers and other adjustments can be a cost-effective way to minimize expected distortions. Starting extended RAPs with smaller, pilot exercises can enable lessons to be learnt to improve the design and implementation of larger exercises (both of which can be further refined over time).

Reporting of PIT gaps estimated from RAPs should be transparent and realistic, recognizing both their limitations and strengths. Where the design of the RAP limits its coverage—for example by not including non-filers in the sample or limiting auditors' examinations to specified risks –the estimated PIT gap will most likely have the same limitations, and be biased downward unless appropriate adjustments are made. And all tax gap estimates are broad point estimates of the level of the gap, though a consistent methodology can ensure they provide robust trend metrics. But a good RAP, comprehensively analyzed and reported, provides an unbiased, detailed analysis of actual taxpayer compliance behavior that can be used to improve compliance risk management and increase revenue yield. And it can improve transparency and governance, providing management and external stakeholders with assurance that case selection of audits does not just rely on old, perhaps biased, perceptions of risk.

APPENDIX 1. RANDOM SAMPLE DESIGNS

A. Simple Random Samples

In a simple random sample, all members of the population (i.e., sample units) have an equal probability of selection. This should produce a sample that is representative of the full population, so results can be scaled by the ratio of sample to population to quantify the tax gap and non-compliance rate in the full population.

Simple random sampling is best suited to homogenous populations. Where there is heterogeneity between population segments, representativeness and sufficient sample sizes of particular subgroups of interest is not guaranteed as few sample units may be selected from those segments of the population as a consequence of random selection. This is particularly undesirable where estimates for subgroups are required.

In the case of random audits, a simple random sample design will result in larger number of observations without non-compliance. This is less useful for analysis when exploring compliance within higher risk segments and will also mean the program yields less revenue.

Sample units can be drawn randomly or systematically depending on whether the data is ordered. At its simplest, random sampling entails drawing each sampling unit (i.e., taxpayer or tax return) at random from the total population, generally represented in the form of a list or registry. In contrast, a systematic sample is drawn by selecting sampling units from a list of taxpayers (or tax returns) at fixed intervals that are calculated to achieve the total sample size required. In this approach, the first case must be selected randomly, to assure equal selection probability of sampling units. A systematic sample can be biased where there is some ordering of the population list by characteristics that fall within patterns that are then systematically either over or under selected. If data is ordered in this way, this ordering effect has to be removed or a sample selected randomly.

An advantage of systematic sampling is it can also be used in the absence of a population list, (e.g., transaction data drawn from a live process). Selecting from a live process can be problematic for planning if the population size is not known and there are restrictions on the final sample size or time period for sampling, as bias could be introduced if selection stops before the full population is covered. This may be especially relevant if considering selecting taxpayers from a live process where timing could correlate with compliance rates, for example the compliance of those filing close to or past a deadline may vary from those who file early or on time.

B. Stratified Radom Samples

Within a stratified random sample, the population is divided into segments (strata) and the sample is drawn with equal selection probabilities within strata and different selection probabilities between strata. The sub-samples for each stratum are simple random samples for the corresponding stratum of the population that are aggregated to produce results for the total population. This means sample units within the population are subject to unequal selection probabilities, but importantly the overall result remains unbiased so long as the results within each stratum are weighted appropriately during the scaling up process (see Box 6).

Stratified random sampling is generally preferred by analysts as a means of increasing the efficiency of a sample, increasing accuracy of the results without increasing costs. The amount of efficiency gains will depend on population and sub-population characteristics, and on the detailed sample design. The choice of strata and their respective sampling rates therefore requires some knowledge of the population and the likely variance between and within population segments. Where individual subgroups are of particular interest (e.g., high-risk taxpayers), the sample size for strata can be chosen to ensure the robustness of the results for the stratum of interest.

In a random audit program, stratification allows over-sampling of high-risk strata. Over selecting high-risk taxpayers will increase higher hit-rates and yield, which is helpful when considering the cost-benefits of the random audit program. Such an approach also likely increases the efficiency of the measurement exercise by boosting estimation precision where there is likely high levels of losses and a lot of variance within specific strata, without requiring larger sample volumes overall. How much particular strata should be over sampled can be determined using various optimum allocation methods.[26]

Under-sampling low-risk taxpayers might produce unusable estimates for this group. In contrast to over-sampling high risk strata, it can be tempting to under-sample low-risk segments or even not sample them at all. Although this can improve the apparent efficiency of the random audit program, it does undermine one of the major benefits of the approach which is not just to estimate the tax gap, but to test and improve existing risk knowledge. A small sample with a low degree of non-compliance will more likely result in too few observations of non-compliance to produce a robust assessment of the risk in that segment of the population. Care must be taken to assure there are sufficient sample numbers for all sub-groups of interest to produce acceptably robust estimates.

BOX 6. Weighting Data

Where cases are selected with an unequal chance of inclusion in a sample, for example, in stratified samples where a particular stratum is over-selected, the sample is not representative of the population so adjustments must be made at the point of analysis. This adjustment is made in the form of weighting data. A weight is a variable added to sample data and is used to adjust how much influence a specific case or stratum of cases has in analysis so that final results are representative of the population. In tax gap analysis, the average value of the gap for each stratum can be multiplied by the proportion of the population that the stratum represents.

C. Single or Multistage Cluster Samples

This form of sample design is more typical in the field of social research surveys and is less likely to be appropriate for random audit program design. Within a cluster sample, the population is first divided into clusters, which are selected with an equal selection probability, then sample units from within the selected clusters are selected with an equal selection probability.

---

[26]   For example, Neyman allocation. See Neyman (1934).

This approach is usually adopted as a practical means of drawing a representative sample when a whole population list is unavailable, but a grouping of the population is known. It is also a way to reduce the sampling costs, for example, by sampling in just a few geographical areas.

A cluster sample design decreases the precision of estimates; therefore, this design is not recommended. The increased uncertainty is presented in results as a 'design effect', a factor that increases standard errors. RAs will typically have a full population list and adequate coverage to access taxpayers in all regions. Where budgets are restricted, a cluster sample design could limit the number of regions in the RAP, but compliance rates can vary by geographical location, so good prior knowledge of the population is required to avoid selection bias.

D. Other Sample Designs

The RAP should employ a probabilistic sampling strategy to produce sound tax gap estimates. Where only specific taxes, taxpayer segments, or risks are in scope of the RAP rather than the whole population—such as international or 'offshore' risks, or certain sectors—it is important to remember that the gap estimate will only represent the population from which the random sample was taken.

APPENDIX 2. NONDETECTION MULTIPLIERS

Non-detection multipliers are used to adjust auditors' assessments of undeclared liabilities to take account of non-compliance that auditors were unable to detect. There are a variety of ways of deriving a non-detection multiplier, including:

a. **Detection controlled estimation.** This is an econometric approach which addresses the varying abilities of individual auditors to detect liabilities, deriving an uplift to RAP results by scaling up what was detected by each auditor to what the "best" auditor would have found in the audit.[27] This approach can provide a multiplier specific to each line item on a tax return and for each auditor based on their detection rates. The data requirements are high. Each auditor must have completed at least 15 cases covering broadly the same line items. Characteristics of auditors (e.g., grade and experience) and specific types of non-compliance are required as explanatory variables in the model.[28] It also only scales up results to the observed capability of the highest performing auditor, which still may not equate with full detection.

b. **Comparing auditor detected liabilities to third party information.[29]** Audits must first be conducted without access to third party information then compared retrospectively to third party data on actual income sources.[30] This will not provide a meaningful non-detection multiplier if auditors already have access to all or most of the available data. Clearly, it will also not provide an indication of the magnitude of non-detection for items where there is no additional data to cross reference, which may include areas of high risk—such as income received in cash.

c. **Secondary case review by expert auditors.** Audits are passed to a separate group of experts, for example, the most experienced auditors, to review where non-detection may have occurred (e.g., items that were not reviewed exhaustively). If cases are passed for review when they are still open, additional checks can be conducted and the difference between the initial revenue identified by the first auditor can be compared to the second. This technique will require additional resource investment and is only as powerful as the abilities of the auditors conducting the secondary review.

d. **Expert judgement.** In this case a panel of experts, for example, experienced auditors, estimate how much tax goes undetected in an audit. This can be explored through a structured series of questions detailing specific scenarios that could contribute to non-detection (e.g., lack of cooperation from the taxpayer, lack of third party data). A Delphi technique can be adopted,[31] whereby respondents provide their independent assessments through a series of rounds, then their responses are pooled to reach a consensus view of the multiplier in the final rounds. A drawback of this approach is there is unlikely to be a way to validate the

---

[27] Erard and Feinstein (2011).

[28] The Internal Revenue Service (IRS) pioneered the use of detection-controlled estimation (DCE). The analysis pooled multiple years (2008-2013) of the random audit program (National Research Program), which has a sample size of approximately 13,000 returns each year, in order to derive estimates using DCE. Internal Revenue Service (2019).

[29] Internal Revenue Service (1988).

[30] This is the approach the Internal Revenue Service used to develop multipliers for their estimates prior to the introduction of detection-controlled estimation. It was based on data from the late 1970s and early 1980s. It may not be as suitable as a method for contemporary tax gap studies now that the use of third-party information is part of business as usual operations within an RA. See Internal Revenue Service (1988).

[31] See HM Revenue and Customs (2020) for a discussion regarding approaches to non-detection multiplier estimation in the UK.

final outcome, and the results will be contingent upon the specific experts on the panel.

e. Adopting multipliers calculated by others. Some RAs have chosen to adopt multipliers calculated by other RAs or experts, or the mid-point of international ranges. This suffers the limitation of possible unrepresentativeness of external multipliers to the country-specific PIT and audit regimes, and specific population under consideration. However, it has the advantage of permitting an RA with limited resources to estimate a theoretically comprehensive tax gap in the absence of an alternative.

The use of appropriate multipliers can reduce the downward bias in gap estimates from non-detection but adds an unknown degree of error to estimates. It is generally almost impossible to validate the accuracy of the final multiplier. If an RA chooses not to employ a non-detection multiplier, the gap can be reported as subject to this limitation, and therefore interpreted as a lower bound or conservative estimate, or defined as the detectable PIT gap.

APPENDIX 3. SAMPLE SIZE EXAMPLE TABLE

A confidence interval sets out a range which has a known and controlled probability (generally 95 percent) to contain the true population value. The margin of error describes half the width of confidence interval in percentage points.

It should be kept in mind that this table describes the margin of error around the estimate of the *rate* of non-compliance, not the *value* of the gap. When considering sample size, analysts should consider what the possible number of yielding cases would be from which to estimate the gap based on possible rates of non-compliance and consider if the sample size would provide a sufficient degree of accuracy. A pilot study based on a smaller initial sample can help provide an indication of the hit-rate and variance that can help guide this decision. In the absence of a pilot study, an appropriate sample size can be estimated assuming the worst-case scenario of incidence[32] (50 percent) and making assumptions regarding the distribution of values for estimating the confidence interval around the gap. This could be informed by drawing on known values and variance from risk-based audits.

Table 2. Sample Sizes by Margin of Error for Different Non-Compliance Rates

Rate of non-compliance		Margin of error in percentage points		
		+/−10%	+/−5%	+/−3%
'Worst Case'	50%	95	385	1065
High	40%	115	370	1025
Medium	25%	70	290	800
Low	10%	35	140	380

Sample sizes rounded to nearest five. Assumes a large population size.

---

[32] Purely from the perspective of the calculated margin of error.

APPENDIX 4. PROJECT MANAGEMENT CHECKLIST FOR A RANDOM AUDIT PROGRAM (RAP)

Planning

Project Management

- Establish working group of analysts, auditors, and other specialists to implement the RAP.

- Establish senior-level oversight of the RAP via a steering group or other means.

Design

- Analyze sample and sub-sample sizes options, their expected accuracy and costs, taking into account stratification options, e.g., taxpayer segments or current risk profiles. Remember:

    o Simple compliance gap estimates should balance costs vs expected accuracy.
    o Risk analysis and profiling need larger samples and more detailed analysis.
    o If the RAP is to increase deterrence effects, sample size needs to be significant.

- Agree audit and analytical resources required (time, staff, other costs).

- Agree case selection, management and reporting requirements, with guidance for auditors.

- Prepare reporting and data capture templates for auditors and agree quality assurance.

- Agree timeline and project plan for RAP and reporting estimates, including milestones.

Scope and Coverage

- Agree non-compliance risks that will not be reliably captured in RAP (e.g., complex frauds).

- Identify taxpayers that are either not in the sample frame or not cost-effective to sample.

- Clarify coverage of expected PIT gap estimate, and how to cover missing areas.

Implementation

Sample Selection

- Sample frame is usually the taxpayer registry, cleaned and up to date.

- For stratified samples, the registry data must include demographic and/or risk profiling data.

- Select cases and share with audit teams, including instructions and templates for RAP.

Audits

- Auditors conduct comprehensive audits of sample cases.

- Provide help desk or other support for auditors to resolve any queries.

- Monitor progress of audits and record any exceptions.

- Quality assure audits, and especially data recording.

Analysis

- Extract data on pre-agreed date and conduct analysis of results.

- Check data for any recording issues e.g., duplicates and open cases. Identify and treat outliers.

- If required, forecast the value of open cases, yield and unrecoverable debt.

- Apply non-detection multiplier and adjust results for unaudited cases.

- Quality assure the analysis and interpret the results.

Reporting

- Follow pre-agreed publication plan to publish gap estimates and associated analysis.

- Report PIT gap as a percent of potential collections, with appropriate context and caveats.

- Follow good practice on transparency, e.g., published methodology and data protection.

GLOSSARY[1]

TERM	DEFINITION
Bias	The tendency of a measurement process to over- or under-estimate the value of a population parameter. It is an effect which deprives a statistical result of representativeness by systematically distorting it, as distinct from a random error which may distort on any one occasion but balances out on the average.
Bottom-up estimation	A bottom-up approach begins with taxpayer-level data and scales up non-compliance identified to the population level to identify the total gap.
Compliance gap	The difference between actual and potential collections, given the existing policy framework; tax lost through non-compliance.
Confidence (statistical)	The degree of certainty of an estimate. In sample statistics, confidence is typically set at 95%, which means there is a 95% chance that the actual population parameter or true value lies within the confidence interval.
Confidence interval	A confidence interval sets out a range which has a known and controlled probability (generally 95%) to contain the true value.
Margin of error	A measure of precision, it describes half the width of confidence interval in percentage points.
National accounts	A system of accounts used to measure countries' economic activity. National accounts should include the non-observed economy.
Non-sampling error	Bias in survey estimates arising from errors in design and implementation; refers to accuracy or validity of an estimate as opposed to its reliability or precision.
Outlier	An outlier is a data value that lies in the tail of the statistical distribution of a set of data values.
Parameter	Parameters are unknown, quantitative measures (e.g., total revenue) for the entire population.
Policy gap	The difference between potential collections under the existing policy framework and potential collections under some normative policy framework.
Precision	The property of the set of measurements having small random error of estimation.
Probability sample	A probability sample is a sample selected by a method based on the theory of probability (random process), that is, by a method involving knowledge of the likelihood of any unit being selected.
Probability or random sampling	Selection methodology whereby each population unit (person, household, etc.) has known, non-zero chance of inclusion in the sample. It is the only general method known which can provide a measure of precision of the estimate. Sometimes the term random sampling is used in the sense of probability sampling.
RA-GAP	Revenue Administration Gap Analysis Program: a technical assistance program to estimate tax gaps for member countries, delivered by the Revenue Administration division in the Fiscal Affair Department (FAD) of the International Monetary Fund (IMF).
Sample frame	Set of materials from which sample is actually selected, such as a list of taxpayers.
Sample size	Number of sample units selected.
Sample unit	The individual element that will be selected when drawing a sample. For example, an individual taxpayer or employer.
Standard error	A measure of the precision of the estimate found in the specific sample. It is calculated as the positive square root of the variance of the sampling distribution of a statistic.
Strata	A subset of the population.
Stratification	Stratification consists of dividing the population into subsets (called strata) within each of which an independent sample is selected.
Systematic sampling	Selection from a list, using a random start and a predetermined selection interval, successively applied.

---

[1] See the OECD glossary of statistical terms for further information. Available at https://stats.oecd.org/glossary/index.htm.

TERM	DEFINITION
Target population	Definition of the population intended to be covered by the study.
Tax base	The sum value of economic activities that generate tax revenues.
Top-down estimation	A top-down methodology uses statistical data to estimate a reference potential revenue value, tax administration data to determine the actual revenue value, and then evaluates the difference between the two to estimate the tax gap.
Weight	An inflation factor applied against raw data.

REFERENCES

Australian Tax Office, 2020, "High wealth income tax gap: Trends and latest findings", Canberra, Australian Tax Office.

Bazzoli, M., Di Caro, P., Figari, F., Fiorio, C.V., Manzo, M., 2020, "Size, heterogeneity and distributional effects of self-employment income tax evasion in Italy". EUROMOD Working Paper EM18/20, Institute for Social and Economic Research, University of Essex.

Biber, Edmund, 2010, "Revenue Administration: Taxpayer Audit— Development of Effective Plans", IMF Technical Notes and Manuals 10/03, Washington DC, International Monetary Fund.

Bloomquist, Kim Michael, Hamilton, Stuart, and Pope, Jeffrey, 2014, "Estimating Corporation Income Tax Under Reporting Using Extreme Values from Operational Audit Data," Fiscal Studies vol. 35, no. 4, pp. 401–419.

Breusch, Trevor, 2016, "Estimating the Underground Economy Using MIMIC Models," Journal of Tax Administration, 2:1, (Apr., 2016) pp. 41-72.

Canada Revenue Agency, 2018, "International Tax Gap and Compliance Results for the Federal Personal Income Tax System," Ottawa, Canada Revenue Agency.

Canada Revenue Agency, 2019, "Tax Gap and Compliance Results for the Federal Corporate Income Tax System," Ottawa, Canada Revenue Agency.

Ceriani, L., Figari, F., Fiorio, C.V., 2020, "EUROMOD Report. Italy 2014-2017". Technical Report, Institute for Social and Economic Research, University of Essex.

Dellaportas, Petros, Ioannidis, Evangelos, and Kotsogiannis, Christos, 2019, "Sample-size determination for risk-based tax auditing," Tax Administration Research Centre (TARC) Discussion Paper 026, TARC, Exeter.

Ministero dell'Economia e delle Finanze, 2019, "Relazione Sull'economia Non Osservata e Sull'evasione Fiscale e Contributiva: Anno 2019", Rome, Ministero dell'Economia e delle Finanze.

Doyle, Melanie, Lepanjuuri, Katriina, and Toomse-Smith, Mari, 2017, "The Hidden Economy in Great Britain," HM Revenue and Customs Research Report 478, London, HM Revenue and Customs.

Engström, Per and Hagen, Johannes, 2015, "Income underreporting among the self-employed: a permanent income approach," Working Paper 2015:2, Uppsala, Department of Economics, Uppsala Center for Fiscal Studies.

Erard, Brian and Feinstein, Jonathan, 2011, "The Individual Income Reporting Gap: What We See and What We Don't," Washington DC, Internal Revenue Service – Tax Policy Centre Research Conference.

European Commission, 2018, "The Concept of Tax Gaps: Corporate Income Tax Gap Estimation Methodologies," Working Paper No 73 – 2018, FISCALIS Tax Gap Project Group, Brussels, European Commission.

Gallucci, Marta, Pansini, Rosaria Vega, Pisani, Stefano, 2020, "Direct Taxes Gap Estimates: Methodology and Preliminary Results", Discussion Topics No. 2/2020, Rome, Agenzia delle Entrate.

Gonzalez Cabral, Ana Cinta, Kotsogiannis, Christos, and Myles, Gareth, 2019, "Self-Employment Income Gap in Great Britain: How Much and Who?", CESifo Economic Studies, 2019, 84–107.

Heckman, James, 1979, "Sample Selection Bias as a Specification Error," Econometrica, Vol. 47, No. 1 (Jan., 1979), pp. 153-161.

HM Revenue and Customs, 2020, "Measuring Tax Gaps: 2020 Edition", London, HM Revenue and Customs.

HM Revenue and Customs, 2010, "Code of Practice for Official Statistics: HMRC policy on Revisions", London, HM Revenue and Customs.

Hurst, Erik, Li, Geng, and Pugsley, Benjamin, 2010, "Are Household Surveys Like Tax Forms: Evidence from Income Underreporting of the Self-Employed", NBER Working Paper No. 16527, Cambridge MA, National Bureau of Economic Research.

Hutton, Eric, 2017, "The Revenue Administration – Gap Analysis Program: Model and Methodology for Value-Added Tax Gap Estimation," IMF Technical Notes and Manuals 17/04, Washington DC, International Monetary Fund.

Internal Revenue Service Research, Analysis & Statistics, 2019, "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011-2013," Publication 1415 (Rev. 9-2019), Washington DC, Internal Revenue Service.

Internal Revenue Service, 1988, "Income Tax Compliance Research: Gross Tax Gap Estimates and Projections for 1973-1992: Supporting Appendices to Publication 7285". Washington, DC, Internal Revenue Service.

International Monetary Fund, 2015, "Current Challenges in Revenue Mobilization: Improving Tax Compliance", IMF Staff Report, Washington DC, International Monetary Fund.

International Monetary Fund, 2019, "ISORA 2016: Understanding Revenue Administration", IMF Departmental Paper, Washington DC, International Monetary Fund

Neyman, Jerzy, 1934, "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," Journal of the Royal Statistical Society, Vol. 97, No. 4 (1934), pp. 558-625.

Organisation for Economic Co-operation and Development, 2002, "Measuring the Non-Observed Economy: A Handbook", Paris, OECD Publishing.

Organisation for Economic Co-operation and Development, 2019, "Tax Administration 2019: Comparative Information on OECD and Other Advanced and Emerging Economies," Paris, OECD Publishing.

Pecho Trigueros, Miguel, Pelaez Longinotti, Fernando, and Sánchez Vecorena, Jorge, 2012, "Estimating Tax Noncompliance in Latin America: 2000-2010," Tax Studies and Research Directorate Working Paper N° 3 – 2012, Panama, Inter American Center of Tax Administrations (CIAT).

Pissarides, Christopher A, Weber, Guglielmo, 1989, "An Expenditure-Based Estimate of Britain's Black Economy," Journal of Public Economics 39 (1989) pp.17-32.

Quiros-Romero, Gabriel, Alexander, Thomas F, Ribarsky, Jennifer, 2021, "Measuring the Informal Economy," IMF Policy Paper No. 2021/002, Washington DC, International Monetary Fund.

Rivas, Lisbeth, and Crowley, Joe, 2018, "Using Administrative Data to Enhance Policymaking in Developing Countries: Tax Data and the National Accounts," IMF Working Paper 18/175, Washington DC, International Monetary Fund.

Rubin, Marcus, 2011, "The practicality of a top down approach to the direct tax gap," in Plumley, Alan ed. "Recent Research on Tax Administration and Compliance Selected Papers Given at the 2011 IRS-TPC Research Conference," Washington, DC, pp. 109-127.

Schneider, Friedrich, Buehn, Andreas, and Montenegro, Claudio E, 2010, "Shadow Economies All over the World: New Estimates for 162 Countries from 1999 to 2007," International Economic Journal, 24: 4, (Dec., 2010) pp. 443-461.

Thackray, Mick, and Alexova, Martina, 2017, "The Revenue Administration – Gap Analysis Program: An Analytical Framework for Excise Gap Estimation," IMF Technical Notes and Manuals 17/05, Washington DC, International Monetary Fund.

Ueda, Junji, 2018, "Estimating the Corporate Income Tax Gap: The RA-GAP Methodology", IMF Technical Notes and Manuals 18/02, Washington DC, International Monetary Fund.

Van Buuren, Stef, 2018, "Flexible Imputation of Missing Data, Second Edition", Boca Raton FL, CRC Press.