Building Towards **Outcome Measurement in Tax Administration**

A guide produced by members of the Fiscalis Risk Management Platform, Project Group on identifying new and innovative methods to measure outcomes in tax administration (Financial Code: FPG/084)

June 2020

Index

1. Introduction

1.1. Background

As society changes, Tax Administrations must continuously adapt the management of taxes (including tax collection). Nowadays the emergence of digitalisation and innovative technologies leads to disruption in the economy and to new business models. These changes offer opportunities for citizens, businesses and governments but also bring new risks for the management of taxes. In the last decade tax administrations have developed more evidence-based compliance risk management strategies to be able to deal with the changing environment in which they operate. Above all, the spectacular growth of available data, the more sophisticated methods of analysing the data and the need to allocate resources more efficiently, have been important drivers for these strategies.

Such strategies encompass: [1]

- shifting from reactive to more proactive approaches, creating an environment in which taxpayers are prevented from making mistakes or committing evasion;
- building of trust and seeking collaboration with taxpayers and third parties to manage compliance (risks);
- influencing taxpayer behaviour by basing the tax administrations' services and enforcement activities on the root causes of taxpayer behaviour, applying an appropriate mixture of preventive and repressive compliance actions;
- taking advantage of the possibilities of digitalisation and new technology to handle large volumes of data to make processes more efficient, to make risk analysis more accurate and to better manage compliance (risks).

Part of changing societies is that modern tax administrations are often held more accountable for what they are doing, why they are doing it and how. They must be able to explain how their actions[2] contribute to their strategic goals. In recent years tax administrations and their stakeholders (society, politicians) focused primarily on the resources spend (input) and the amount of work done (output) when discussing the workings of the tax administrations. However, these metrics cannot confirm if the actions deployed also have led to the expected change in the behaviour of taxpayers (compliance). Whether or not the intended goals are being achieved is a crucial question to be answered.

Outcome measurement can help answer that question. It is defined as a part of the evaluation that describes if and to what extent the desired objectives of a tax administration are achieved, caused by the actions of the tax administration.

This is necessary to enhance the application of a compliance risk management strategy and to discuss the results of the work done in an adequate way.

Outcome measurement however proves to be a challenge for most tax administrations because they are struggling with the concept of putting outcome measurement into practice. Therefore, the Fiscalis Risk Management Platform decided to start a Project group (FPG 084) to discuss outcome measurement within the context of Compliance Risk Management and explore the topic with a view to creating hands on guidance for tax administrations.[3]

1.2. Motivation

This report aims to offer guidance to tax administrations to start on the path to outcome measurement.

The report discusses the place of outcome measurement within a coherent evidence-based compliance risk management strategy, examines the strengths, weaknesses, opportunities and threats (SWOT) of outcome measurement and provides practical guidance on how to start. The report contains numerous examples of the measurement of outcomes on various levels within a tax administration (strategic, tactical and operational).

[1] Tax Administration 2019 (OECD)
[2] By "actions" we mean all activities, instruments, interventions, measures, programs, etc., or, in other words, everything an administration undertakes in order to carry out its strategy.
[3] The Project group is chaired by Ireland and the Netherlands and comprised of delegates of Austria, Belgium, Bulgaria, Finland, Ireland, Italy, Lithuania, the Netherlands, Spain and Sweden (see also annex 6.3).

The report also addresses the use of the results of evaluation and the way a tax administration could explain internally and externally what it has done, and why and how it has done it. The value of "story telling" for successful understanding of outcomes is emphasised.

The innovative value of the report is reflected in the fact that the report delivers:
- a 'standard' for tax administrations for outcome measurement in practice;
- the link between underlying academic literature and outcome measurement in practice;
- new practical examples to cover 'blind spots'.

All in all, this report provides practical guidance and knowledge that should be of great value to all the EU's tax administrations.

1.3. Content and Reading Guidance

The focus of the report is on providing material of direct application and use to tax administration staff responsible for outcome measurement: practical examples, suggestions for first steps to building towards outcome measurement and guidance on storytelling about outcomes are the primary deliverables for such readers.

Following the Introduction *(Chapter 1)*, the report is structured as follows:

Chapter 2, Introduction to Outcome Measurement in Compliance Risk Management, contains a general introduction on the topic of measuring outcomes including defining what is meant by 'outcomes' in this context. The chapter explains how outcome measurement fits within the broader goals and approach advocated for Compliance Risk Management by the Platform.

Chapter 3, Process of Outcome Measurement in Practice, explains in detail how to prepare, design, implement and communicate the delivery of outcome measurement. This includes two extensive examples based upon which all steps are illustrated.

Chapter 4, Examples and literature, gives an overview of a selection of activities of a tax administration within a Compliance Risk Management context that could be the subject of outcome measurement. Next these activities are split into proactive and reactive actions and finally these are linked to an extensive literature database and to a database of practical examples of tax administrations. Where possible the whole report is supported by references to examples or literature showing the use of outcome measurement (or the measurement methodology approach).

Chapter 5, Executive summary and suggestions, is the concluding chapter, in which the most important content is highlighted and some suggestions for the successful implementation of 'outcome measurement' within a Compliance Risk Management strategy are proposed.

In addition to these chapters, in the form of appendices or databases, further resources are produced for the use of staff of tax administrations working on outcomes: a comprehensive literature list, a brief summary of relevant evaluation methodologies and background documents, and a list of examples of outcome measurement by tax administrations.

Relevant documents are mentioned in footnotes; clicking on the name of the document will give access to the original document (when available).

This report aims to be useful and accessible to anyone working in tax administration, from policy makers, senior management, analysts to case workers, who has an interest in evaluation, measuring outcomes and ultimately the effectiveness of tax administration.

Reading guidance

Chapter	Reader
Chapter 1 Introduction	Policymakers, management, CRM experts, analysts, caseworkers
Chapter 2 Introduction to outcome measurement	Policymakers, management, CRM experts, analysts, caseworkers
Chapter 3 Process of outcome measurement	Analysts, caseworkers
Chapter 4 Examples and Literature	CRM experts, analysts, caseworkers
Chapter 5 Conclusions and discussion	Policymakers, management, CRM experts, analysts, caseworkers

2. Introduction to Outcome Measurement in Compliance Risk Management

This section, Introduction to Outcome Measurement in Compliance Risk Management, contains a general introduction to Compliance Risk Management (par. 2.1) and the topic of measuring outcomes including defining what is meant by 'outcomes' in this context. The chapter explains in par. 2.2 how outcome measurement fits within the broader context of evaluation, as one of the stages of Compliance Risk Management. The chapter also discusses the strengths, weaknesses, opportunities and threats of outcome measurement for tax administrations in par. 2.3 and provides general recommendations for embedding outcome measurement within a Compliance Risk Management strategy in par. 2.4. Furthermore, references are summarized in par. 2.5.

2.1. Introduction to Compliance Risk Management

Modern tax administrations meet the challenges arising from rapid changes within society with the concept of Compliance Risk Management. The concept is described in several EU and OECD publications and can be defined as follows:

"Compliance Risk Management (CRM) is a systematic process in which a tax administration makes deliberate choices on treatment instruments that could be used to effectively and efficiently stimulate compliance and prevent non-compliance, based on the knowledge of all taxpayers (behaviour) and related to the available capacity." [4]

The CRM process is represented in Figure A:

***Figure A**: Compliance Risk Management Model*

Explanation:
The objective of Compliance Risk Management is to enable a tax administration to accomplish its strategic objectives by facilitating management to make better decisions.

To be able to do so, the context within which tax administrations and taxpayers operate should be clear. The context can be defined as the combination of all facts or circumstances that surround an organisation including both external factors (e.g. legislation) or internal factors (e.g. resource availability) and which, in the short to medium term, should be considered as a given. The objectives of the organisation describe 'what to achieve' in terms of stated goals (e.g. compliance). The Compliance Risk Management process is based on 5 consecutive steps, which form a learning cycle. The first two steps relate to risk identification, analysis of risks and the behaviour of taxpayers from which the risks stem. The next two relate to treatment

[4] Fiscalis Risk Management Platform: EU Compliance Risk Management Guide for Tax Administrations (2010).

planning i.e. making choices – about (groups of) taxpayers, risks and options for treatment – and the implementation of the treatment. The final step relates to evaluation and learning. All steps revolve around the goals implied by the strategy of the tax administration.

Within this context 'outcome' can be defined as a part of the evaluation step, that describes if and to what extent the desired objectives of a tax administration are achieved, caused by the actions of the tax administration.

2.2. Evaluation

Evaluation is broader than looking at 'outcomes' only. It could be beneficial for tax administrations to look at the quality of action plans (plan evaluation in par. 2.2.1) or the process of actions (process evaluation in par 2.2.2). The choice for the specific evaluation method depends on what a tax administration wants to know or what is possible. All forms of evaluation could contribute to various goals as will be discussed in par. 2.2.3.[5]

2.2.1. Plan (or formative) evaluation

A plan evaluation evaluates the policy or intervention plan, i.e. the plan describing how the administration intends to achieve its goal(s). This type of evaluation answers the question if the underlying plan in which the activity is elaborated fits within the Compliance Risk Management strategy of a tax administration and whether it is reasonable to assume that the (mix of) compliance actions planned, will contribute to reaching the objective set.

Relevant questions to be asked in a plan evaluation are e.g.:

- Does the plan fit within the general framework of Compliance Risk Management?
- Is the plan targeted at the right group of taxpayers?
- Are the underlying assumptions of the plan clear?
- Are the underlying assumptions supported by evidence (e.g. from academic literature, previous experience) (see example 1)?
- Is it plausible that the set-up will lead to the intended results?
- How will success or failure of the plan be assessed?
- Is it useful to carry on with the plan?

[5] Types of evaluation

The timing for carrying out a plan evaluation is before the implementation of the activities as a plan evaluation shows how promising the activities are in reaching the goal(s). But also during the implementation this type of evaluation can be worthwhile to decide if it is feasible to continue the implementation.

The plan evaluation is carried out by having the plan scrutinized by compliance risk management experts and/or behavioural scientists who can assess whether the intended activities fit within the Compliance Risk Management strategy and address the root causes for non-compliance and whether the assumptions on effectively influencing behaviour are justified (see example 1).

Example 1: Plan evaluation

The Netherlands Tax and Customs Administration (NTCA) developed a so-called "tax intermediary approach". The policy-theory supporting this approach assumes that a fiscal intermediary will improve its client's compliance when it has the knowledge, the motivation and client's permission to file correctly.

In the plan evaluation the NTCA checked these underlying assumptions of the "tax intermediaries approach" based on academic literature. Research showed that most actions undertaken in this approach can be linked to these three assumed drivers of compliance improving behaviour. Furthermore, research also found a link between these drivers and tax intermediaries' compliance promoting behaviour.

Figure B shows the described elements of the plan evaluation:

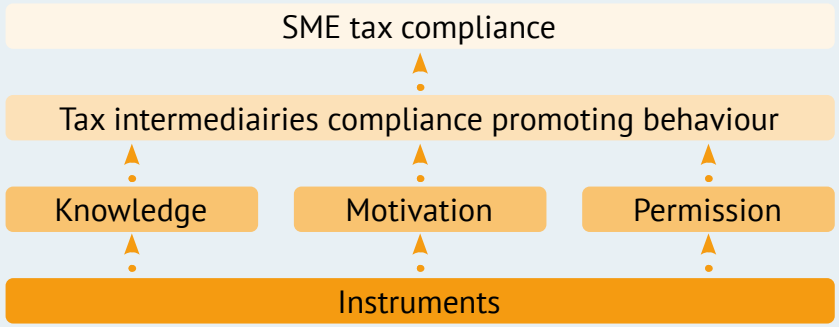


Figure B: *Underlying assumptions plan evaluation*

2.2.2. Process evaluation

A process evaluation focuses on the way the compliance activities are carried out. A process evaluation answers the question how and/or why a plan works (or doesn't work).

Relevant questions to be asked in a process evaluation are:

- Are the actions executed as planned?
- If not, were the right adjustments made to keep actions aligned in a way that they still plausibly lead to the goal set?
- Are all preconditions in practice fulfilled?
- What is the opinion of the staff implementing the activities?
- Are the costs of the activities and the time to finish in adherence with the budget?
- Is the level of cooperation between the parties involved sufficient?

To prevent the process evaluation being based on coincidental evidence, it is advisable to accumulate experience from repeated activities.

A process evaluation can be carried out by analyzing the project documentation and interviewing staff but also taxpayers involved. Interviewing can take place in many different forms to include, semi-structured interviews, group discussions or surveys. Briefing and debriefing of staff before and after the activity may also be a possibility. It is preferable to log the execution of the intervention and to record relevant issues like obstacles, choices made, success factors etc..

2.2.3. Outcome measurement

Outcome measurement is a form of evaluation to determine if a tax administration has reached its strategic goals (e.g. compliance). It also may contribute – in the long term – to answering the question if a tax administration has reached its 'societal' goals, that have led to a change in society (e.g. tax morale) (see example 2). Outcome measurement describes if and to what extent the desired objectives of a tax administration are achieved, caused by the actions of the tax administration.

Example 2: Societal goals

An example of a societal goal is to decrease the widespread tax avoidance/evasion. The implementation of the worldwide standard of Automatic Exchange of Information (AEOI) that contains detailed information about the financial accounts each jurisdiction's taxpayers hold abroad should help to reach that goal. The effectiveness of the working of the AEOI Standard in practice needs to be assessed to conclude if the AEOI indeed had led to less tax avoidance.[6]

Tax administrations can have various reasons to measure outcomes:

1. Learning (to improve its compliance risk management strategy): is the administration progressing towards better compliance rates or not, which activities turn out to be the most effective, what can be improved?
2. Managing (to be able to make educated choices with regard to capacity and instruments): by showing which activities turn out to be more or less effective, the administration can direct its activities towards the most effective ones.
3. Accountability (to show the added value of its activities): the tax administration can demonstrate to parliament/government the effectiveness of their activities in meeting the (compliance) objective(s).
4. Reflecting (to discuss e.g. the feasibility of legislation): outcome measurement can show that progressing towards the compliance objective is beyond its control and for instance the legislator is the only one that can make a difference.

As SMART objectives and a sensible policy theory are prerequisites for outcome measurement, outcome measurement will have a disciplining effect in applying Compliance Risk Management. Tax administrations are 'forced' to be more clear on the goals they want to achieve.[7]

Measuring outcomes as part of a Compliance Risk Management strategy means answering the question whether and to what extent the activities of the administration caused a (positive) change in the behaviour of taxpayers. Chapter 3 elaborates on how this is done in practice, illustrated by examples.

2.3. SWOT analysis on outcome measurement

To support the discussion within a tax administration whether it would be feasible and interesting to implement outcome measurement as part of Compliance Risk Management, a SWOT-analysis is undertaken in which the possible Strengths,

[6] Global Forum on Tax Transparency marks a dramatic shift in the fight against tax evasion with the widespread commencement of the automatic exchange of financial information

[7] SMART is an acronym for being Specific, Measurable, Assignable, Realistic, and Time-related.

Weaknesses, Opportunities and Threats (SWOT) of the organisation in relationship with the implementation of outcome measurement are described. The results of the SWOT analysis are summarized in figure C.

Strengths

- Evidence-based strategy
- Strategy validation
- Clear goal setting
- Improvement in quality of services
- Influencing taxpayer behaviour
- Complete picture/in-depth understanding
- Availability of "big data"

Weaknesses

- No "baseline-measurement"
- Resource and/or skills deficit
- Potential additional costs
- No SMART objectives
- Internal opposition
- Poor data quality and availability
- Lengthy preparation time
- Impediments (legal, practical, ethical etc.)
- Potential absence of short-term results

Opportunities

- Improvements in tax morale and trust
- Shaping changes in policy/legislation
- Moving focus "upstream" (right from the start)
- Technology (data collection & analytics)
- Collaboration with academics

Threats

- Lack of political support
- Reduction in tax administration staff
- Frequent changes in management (policy)
- Changes in the economy
- Changes in legislation
- Reputational risk
- Media focus

Figure C: *SWOT analysis on outcome measurement*

Explanation SWOT analysis:
As outcome measurement presents the tax administration as modern, state of the art, rational and critical towards itself, this can induce trust and credibility towards taxpayers, politics and society. It can stimulate clear goal setting and show tax administration's added value and in doing so, enhance belief in the applied Compliance Risk Management strategy and bolster self-esteem.

Outcome measurement can also be helpful in moving focus upstream; this means investing more in preventative measures before the tax return has been filed, instead of only deploying repressive measures.

'Big data' means extremely large data sets that can be analysed computationally to reveal patterns, trends and associations, especially relating to human behaviour and interactions. This vast amount of data at the disposal of tax administrations and the growing availability of data-analysts and behavioural scientists are very promising developments for a growing use of outcome measurement. However, the quality of the data and the limited number of analytical staff can be a hindrance.

To overcome that hindrance, cooperation with the academic world could be enhanced. Also, making full use of current technology makes it possible to move the Compliance Risk Management strategy more upstream and thus making more use of instruments preventing taxpayers from making mistakes.

'Baseline measurement' is the process of establishing the starting point of any metric from which the improvement or impact of any change measure is calculated. It is used to gauge how effective an improvement or change initiative is. This issue has a rightful place among the weaknesses as often researchers get involved too late at which point it may be no longer possible to measure the baseline.

Another serious challenge for outcome measurement can be the lack of continuity. Often activities are changed or abandoned as a result of new urgencies, whether or not caused by political priorities. Outcome measurement in terms of establishing a causal relationship between the actions of a tax administration and the actual results is not always feasible due to a number of factors to include legal, practical or ethical limitations. The cost effectiveness can also be arguable.

Chapter 3 describes the practical process of carrying out outcome measurement by a tax administration. In this chapter we discuss alternatives to overcome the weaknesses and threats mentioned in the SWOT analysis.

2.4. Implementation challenges

Implementing evaluation calls for choices to be made on various levels in the tax administration. Management has to decide what kind of evaluation (plan, process, outcome) it wants, which level of certainty on the (causal) relationship between the tax administrations' activities and outcomes and thus, what alternatives are desirable (see par. 3.2.3. Step 3: Finding a research design). Also if resources are scarce or the possibilities to execute outcome measurement are limited, management choices have to be made. Therefore it is necessary that a tax administration has a clear vision/policy on evaluation or – more specific – outcome measurement.

Relevant criteria for making choices can be the societal relevance of the objective(s) of actions, the costs involved in the actions, the administrative burden on taxpayer, equality of treatment, ethical issues and the potential generalizability and applicability of the results.

The search for effective actions is a question of behavioural insights, data analysis and experienced staff. This hardly ever leads to one decisive course of action and therefore often needs experimenting. Conducting smaller experiments is a good starting point for implementing outcome measurement. However, it is important to implement outcome measurement as a structured and systematic learning process (within Compliance Risk Management) but not as a stand-alone activity.

Management sometimes is hesitant to implement outcome measurement due to the fact that results are seen as 'right or wrong' instead of 'confirmation of underlying assumptions'. Outcome measurement enables a tax administration to communicate more objectively about the results of its CRM strategy and gives the possibility to communicate more positively or pro-actively – internally or externally – instead of defensively or reactively.

Outcome measurement as part of Compliance Risk Management also draws attention to the currently often still prevailing practice of input and output KPI's. For applying a Compliance Risk Management strategy in an adequate way, a full package of KPI's (input, output, outcome etc.), both quantitative and qualitative, are needed. It is important to have a full narrative ("storytelling") about the broad results of an activity, a project or group of taxpayers, instead of just focusing on one or two KPI's.

A narrative is a better alternative or addition to the current practice of accounting for the tax administration's strategy. A sample narrative might be:

1. What compliance issues did the administration identify?
2. What was the underlying root-cause?
3. How did the administration handle the problems identified?
4. What kind of activities did the administration undertake, what were the resources used, what was the output, what were the side-effects of the activity, what external factors were of influence, what changes in the behaviour of taxpayers were observed (outcome)?
5. How did the administration evaluate the actions, what were the conclusions and the lessons learned?

2.5. References

If you want to know more:

- Compliance Risk Management Guide for tax administrations, Fiscalis Risk Management Group, EU 2010
- An Overview of Evaluation Methodology and Techniques, Fiscalis Risk Management Group, EU 2010 (not published)
- Measures of Tax Compliance Outcomes, a practical guide, OECD 2014
- Types of evaluation
- Process evaluation
- Process evaluation versus outcome evaluation

3. Process of Outcome measurement in practice

The third section, Outcome Measurement in Practice, explains (after an overview in par 3.1) in detail how to prepare and design (in par. 3.2), implement (in par. 3.3) and communicate (par. 3.4.) the delivery of outcome measurement. This includes a complete set of elaborated real case studies and lessons learned from practical experiences. In par. 3.5. references are summarized.

This chapter offers step-by-step guidance on how to proceed with outcome measurement in practice. It gives detailed advice, examples and thought-provoking questions to the professionals implementing outcome measurement, both on the data and policy sides. Understanding of basic statistics is assumed, but the concrete examples give novel ideas on how to establish outcome measurement in the context of tax administrations.

3.1. Introduction

The Project Group has developed a framework for putting outcome measurement into practice within a Compliance Risk Management strategy. This section presents an overview of the suggested steps of the outcome measurement process (figure D).[8]

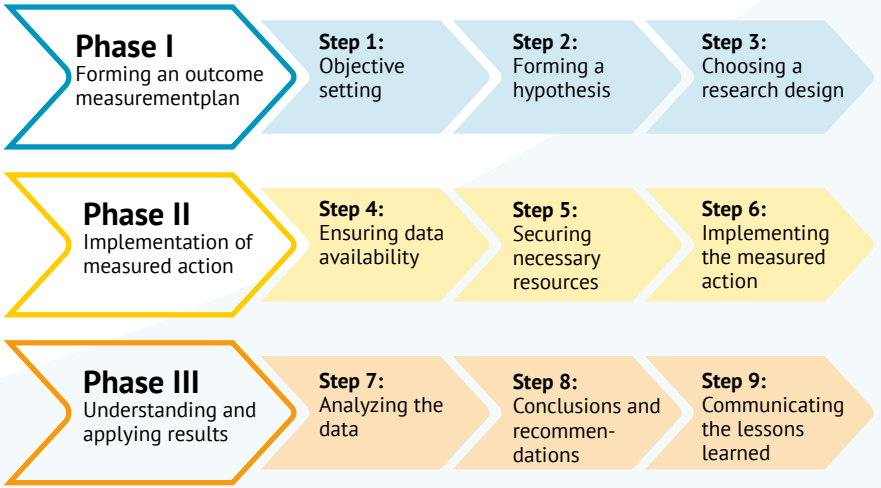


Figure D: *Outcome measurement steps in practice*

The outcome measurement process has been divided into three phases, which in turn consist have three steps each giving a total of nine steps overall. The first phase, Forming an outcome measurement plan, includes steps 1 to 3. The second phase, Implementation of the measured action, includes steps 4 to 6, and the third and final phase, Understanding and applying the results, includes steps 7 to 9. Under each of these steps, the appropriate actions or issues to consider are described.

The steps are laid out with the aim of offering easily applied tips, ideas and instructions. While the steps describe the actual tasks related to outcome measurement, the steps also touch base with the action being evaluated, so to give an idea of the temporal relation to the action.

To provide an idea of how all the steps come together in practice, two running examples – from the Swedish Tax Agency (STA) about preventative audits and Italian Revenue Agency (IRA) about repressive use of tax audits – are used throughout the description of the steps.[9] The complete examples are to be found in the annexes (6.1 and 6.2).

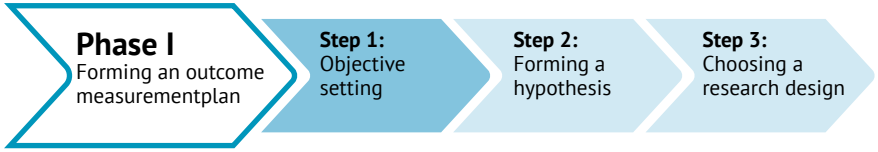
[8] The composition of steps presented here draws from the practical experiences in outcome measurement within the EU member states. While the steps are presented in a certain order, most, in practice, go hand-in-hand with each other and could overlap. They are not set in stone and are adaptable depending on specific circumstances and needs.

[9] The running examples are based upon a practical example of the STA and the IRA which are modified for educational purposes.

3.2. Phase I: Forming an outcome measurement plan

In this paragraph the steps to come to an outcome measurement plan are discussed. An outcome measurement plan contains three important elements, objective setting (step 1), forming hypotheses (step 2) and choosing the research design (step 3).

3.2.1. Step 1: Impulse and objectives for outcome measurement



A tax administration can have various reasons for measuring outcomes as is described in par. 2.2.3, but outcome measurement is usually prompted by a need for information.

How the actual outcome measurement will look is based upon the reason for the outcome measurement and the goals a tax administration wants to achieve. In general a tax administration has goals on various levels: operational goals for the short term (on the level of activities/instruments - see example 3), tactical goals for the mid-term (on the level of specific groups of taxpayers - see example 4) and strategic goals (on the level of the whole population taxpayers - see example 5) for the long-term.

Example 3: Operational goals

On an operation level the Polish Tax Administration with regard to 'reminder letters' wanted to test:
- the effect of different behavioural messages on income tax compliance
- if the method of delivering the letters (regular versus registered mail) had an effect on compliance.

Example 4: Tactical goals

On a tactical level the Austrian Tax Administration wanted to verify the effect of early supervision on the compliance of newly founded firms.

Example 5: Strategic goals

On a strategic level, the Finnish Tax Administration defines its goals as:
- ensuring tax revenue
- fair tax assessment and
- positive taxpayer experience.

Defining clear objectives for the action a tax administration wants to take – including making explicit what the underlying mechanism is – is key to successful outcome measurement, and a natural first step in the process. The objectives must be as clear as possible, as they will determine what will be measured. If the goal and how it links to different levels of operations is not clearly defined, finding the right indicator that accurately describes whether the objective has been reached becomes difficult and the results could focus on the incorrect areas/issues.

To make sure we realise all goals or objectives, and make clear how they are connected on different levels, it can be useful to represent them in an 'objectives tree' as illustrated in Figure E. The objectives tree breaks down the main (higher-level) goal into smaller (lower-level) ones (where the lower-level ones could be considered as the means to achieve the higher-level ones). In general these lower-level goals are more easily measured compared to the higher-level goals. An objectives tree in general needs to be 'constructed'. Policy or strategy papers, project plans, interviews with stakeholders or other such sources may form the basis for its "construction".

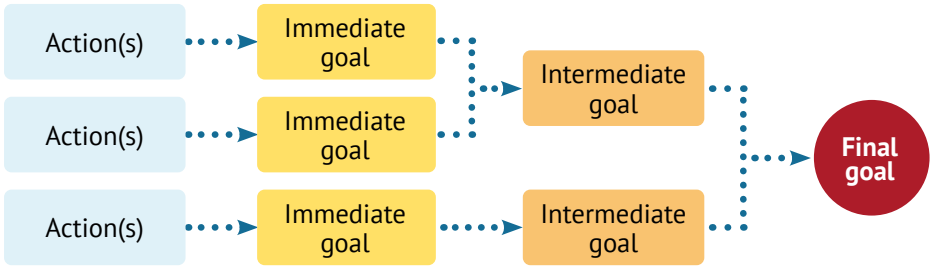


Figure E: *Objectives tree*

A tax administration should be aware of multiple actions towards the same (middle-level) objectives already in existence when planning the policy action. The more activities a tax administration deploys for the same (sub)goal, the more difficult it will be to determine the part of each in the result. It may be that the result would even have been better if only one of the activities had been carried out (see example 6).

Example 6: Multiple activities

The Belgian tax administration started to send reminder letters to taxpayers who did not file a Personal Income Tax return. The number of non-compliant taxpayers included many retired people with only a pension for income.

At the same time, the tax administration started to send proposals of tax bills to those taxpayers for whom all necessary data were already available. This proposal of tax bill could be considered as a filed tax return. Most of the retired people fell into this category.

At the end of the year, the tax administration found that the ratio of filed tax returns had significantly raised. This increase however was not entirely the result of the reminder letters but was largely impacted by the 'tax bill proposals'

TIP: *A good guideline for defining the objectives are the SMART[10] criteria.*

[10] SMART is an acronym for being Specific, Measurable, Assignable, Realistic, and Time-related.

TIP: *Once an objectives tree is finalised, a suggestion is to discuss it with management, project leaders, or other stakeholders to make sure the framework is correctly understood.*

STEP 1 –STA EXAMPLE[11]

In 2010, the Swedish Tax Agency launched a program called "Preventative Audits". In this program, on-site audits were carried out prior to the filing due date for Small and Medium-sized enterprises (SME), within a certain region, who filed late in 2009.
The program is a one-time trial and the purpose is to understand whether it is effective to promote "voluntary compliance" with this new way of working.
The objectives that STA wanted to evaluate are whether preventative audits can lead to:

- Increased correct filing
- Increased filing on-time
- Increased declared incomes/taxes

On a more strategic level, these objectives contribute to an increase in voluntary compliance, as seen in the objective tree (figure F).

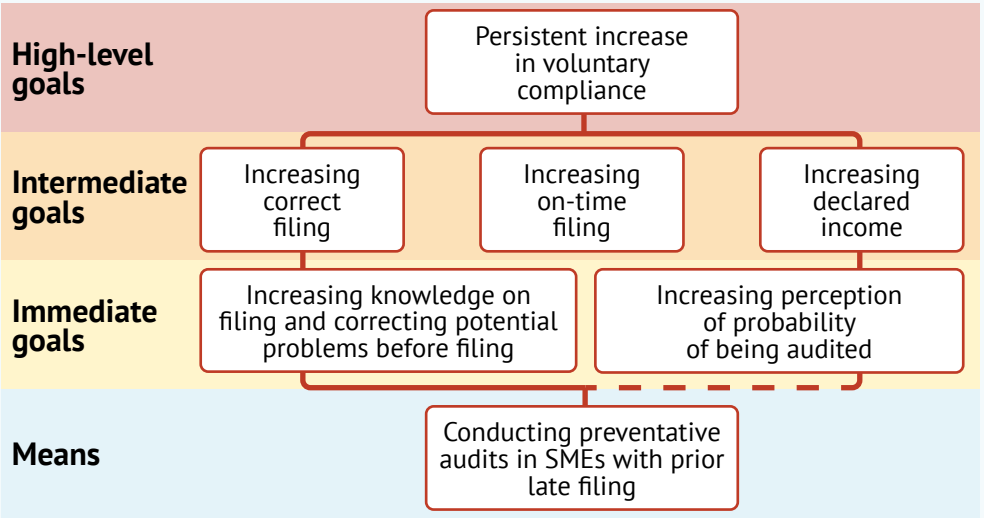


Figure F: *Objectives tree, STEP 1 – STA example*

[11] The running example is based upon a practical example of the Swedish Tax Agency (STA), which is modified for educational purposes.

STEP 1 – IRA EXAMPLE[12]
A tax audit is one of the instruments in the Compliance Risk Management (CRM) strategy applied by Italian Revenue Agency (IRA) to tackle tax evasion and to promote compliance. Each year, IRA launches a risk based audits campaign on self-employed and small businesses (sole proprietors).

With this campaign, in which different types of audits are deployed, IRA aims:
1. to redress non-compliance and collect the revenue involved
2. to increase compliance (correct and complete filing) of audited taxpayers[13]

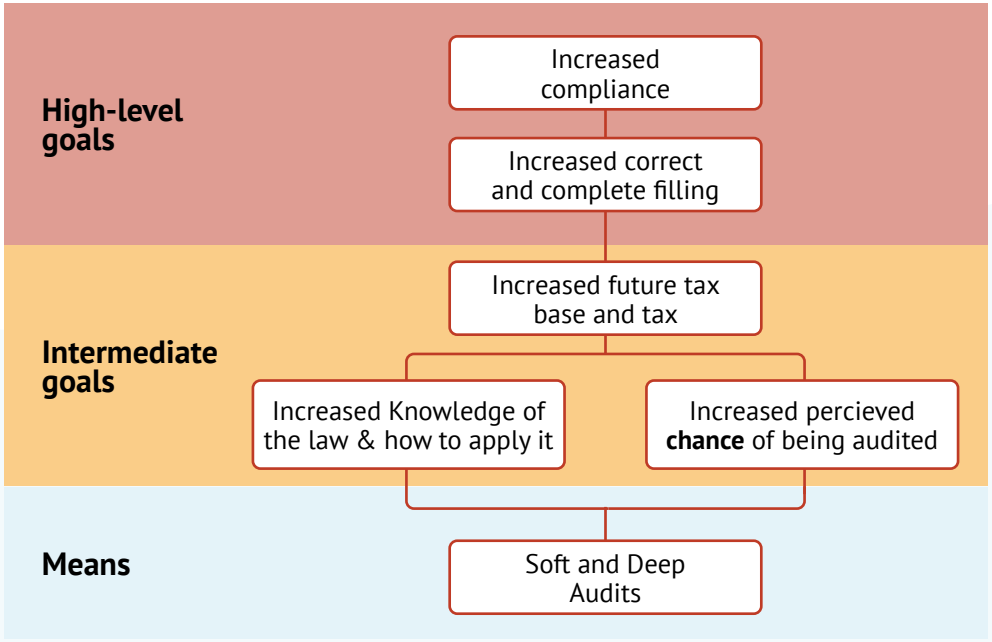
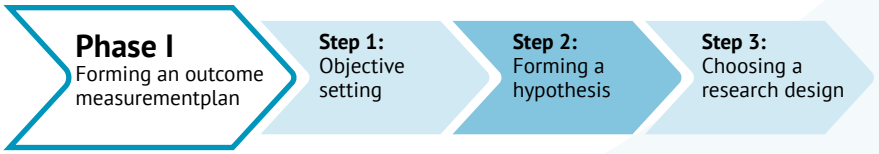


Figure G: *Objectives tree, STEP 1 – IRA example*

[12] The running example is based upon a practical example of the Italian Revenue Agency (IRA), which is modified for educational purposes.
[13] A third goal, an increase in correct and complete filing of non-audited taxpayers, is not treated in this example. The findings can be found in the paper "Tax compliance and the role of the consultant, evidence from an Italian experience" Alfonso Carfora, Elena D'Agosto and Stefano Pisani.

3.2.2. Step 2: Forming a hypothesis



After the desired objectives have been defined (and the objectives tree is constructed – see step 1), the applicable research questions (hypotheses[14]), focussing on the mechanism for the change, need to be established. It is often illustrative to consider the mechanism as a chain or network of factors and events (a so-called effect chain) that follow or are connected to each other, starting from the planned action and ending at the desired outcome.

An *effect chain* (see figure H) is a useful tool for illustrating how the mechanism is expected to work. It draws, among other things, from Program Theory or Program Logic, which is a concept aiming to understand the full set of factors involved in an action or "program", and what kind of impact they will have.[15] Program Theory or Program Logic set out to identify the long-term, medium-term and short-term outcomes, as well as inputs, outputs and assumptions made. They are also useful for breaking down large problems into more manageable pieces.

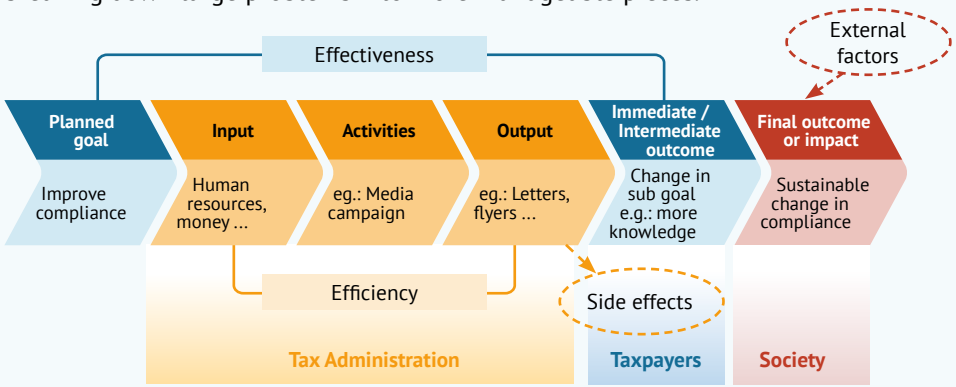


Figure H: *Effect chain*

[14] Hypotheses
[15] Evaluation toolbox

For establishing the hypothesis, all elements of the 'effect chain' need to be considered, such as:

- What is the connection and mechanism between the planned action and the change in outcome?
- What is the chain of cause and effect and what are the links of that chain?
- Are there external factors, such as components in the operating environment or in the tax administration's own functions that may be realized in the intended outcome, not related to the action at hand, e.g. a change in legislation?
- Are there expected side-effects[16] deriving from the action, as in impacts on objects other than the intended outcome?

As prompted by the questions above, external factors and side-effects should also be considered when pinning down the hypothesis. External factors are deriving from outside the tax administration but affecting its outcomes. Examples of external factors are: economic conditions, legislation, demographic developments. Also a tax administration has to take into consideration possible (positive or negative) side effects. Side effects are other effects than the intended ones, deriving from the actions of the tax administration (see example 7).

Example 7: Side-effects

The Swedish Tax Administration launched a campaign to try to increase the use of a digital mail box. Under this initiative, taxpayers were offered a tax settlement before Easter if they accepted the prefilled tax return via the digital mailbox. A side effect could be that a taxpayer may take the offer in order to get an early tax settlement but applies for deductions in a later stage via an appeal procedure, which may increase the administrative cost of the tax administration.

The effect chain also makes clear how the flow of activities is going, starting at the tax administration that carries out activities (and can decide and steer on input and output) to influence the behaviour of taxpayers (outcome), which ultimately must lead to a (long lasting) impact in society.

TIP *It is useful to see if any existing empirical results on the phenomenon are available to help in forming the right hypothesis and action plan. For this purpose, among others, this report includes a comprehensive database of literature on measuring tax compliance and risk management activities.*

TIP *Answering the questions regarding the planned action is important, as there are cases where many potential theoretical channels exist, and the results could be ambiguous in terms of the dominant mechanism.*

STEP 2 – STA EXAMPLE

STA's "Preventative Audits" were expected to have two main effects on tax compliance: a direct effect and an indirect effect. The direct effect comes from the mistakes directly detected and prevented during audits before the filing due date, therefore resulting in additional collected revenue. The indirect effect is related to changes in the behaviour of the audited taxpayer after the auditing and can be denoted as a deterrence effect.

The underlying assumption of obtaining the expected deterrence effect by preventative audits are as follows:

- It provides assistance and education to taxpayers about how to fill in the tax return correctly
- It can increase the perception of the possibility of being audited.

Effect chain of the STA EXAMPLE:

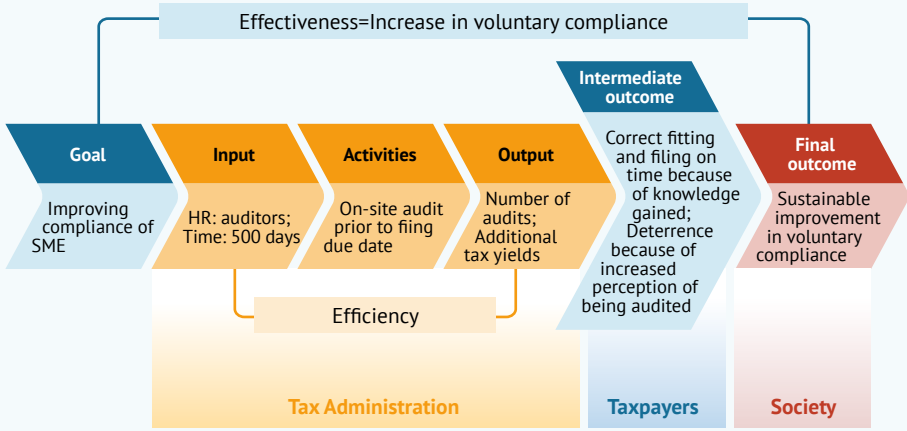


Figure I: *Effect chain, STEP 2 – STA example*

[16] What are externalities?

STEP 2 – IRA EXAMPLE
IRA expects a positive compliance effect because it assumes that during the auditing process the taxpayer learns about the right interpretation and application of fiscal law. Furthermore IRA expects a compliance effect because it assumes that audits increase 'audit-awareness' and with that, the perception of getting "caught".

IRA takes an increase in tax base and tax declared as a proxy of an improvement of compliance (correct and complete filing).

Effect chain IRA EXAMPLE:

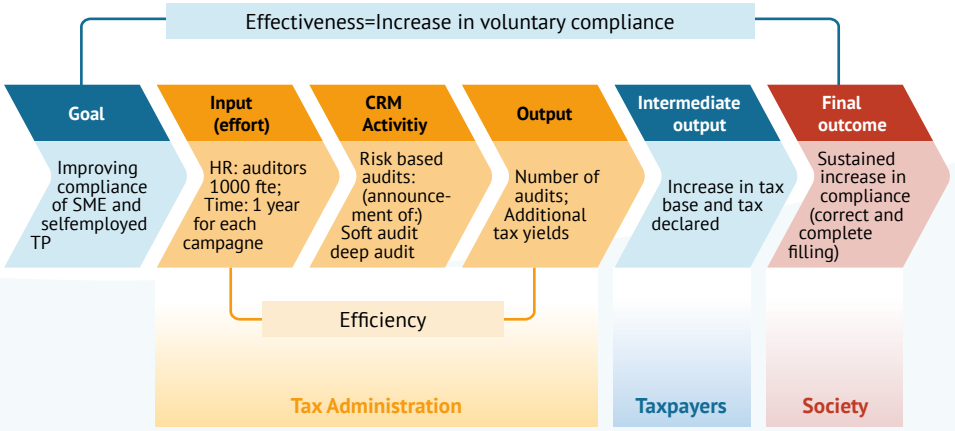
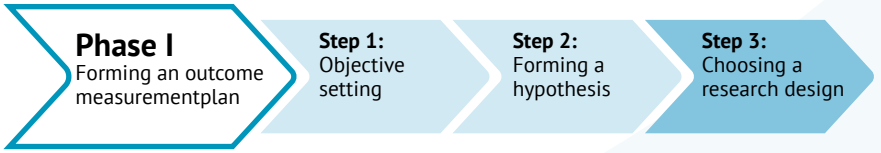


Figure J: *Effect chain, STEP 2 – IRA example*

3.2.3. Step 3: Finding a research design



After forming the hypotheses (in step 2) the next step is to find a research design that will establish if the action of a tax administration has indeed led to achieving the desired goal(s), solving the underlying problem. The Maryland Scientific Methods Scale (MSMS)[17] ranks different evaluation methods based on how robustly causal links between desired goals and actions of the tax administration can be established. The MSMS and its different levels, are presented in figure K.

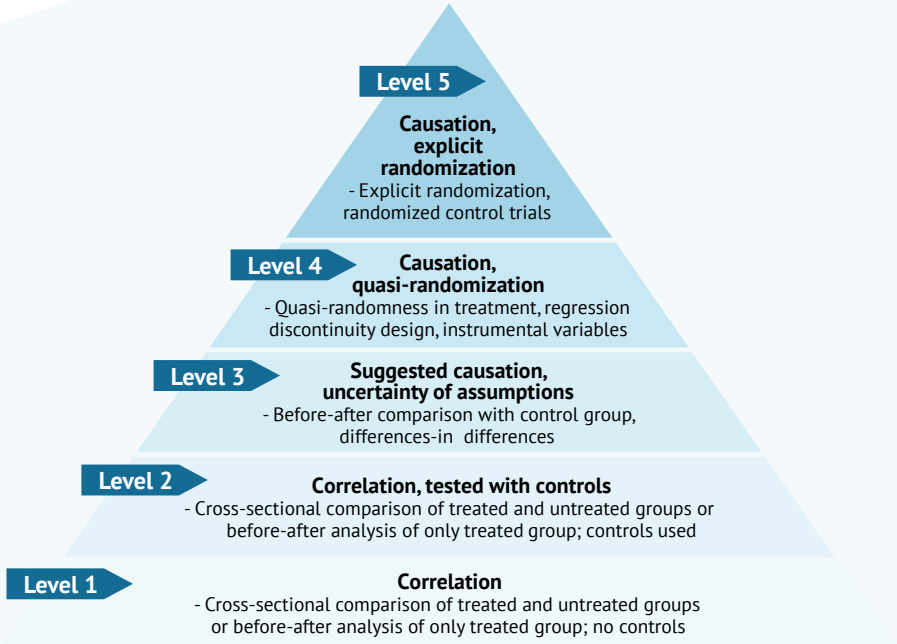


Figure K: *Maryland Scientific Methods Scale (MSMS)*

[17] Maryland Scientific Methods Scale (From Evidence-Based Crime Prevention, P 13-21, 2002, Lawrence W. Sherman, David P. Farrington, et al

The MSMS distinguishes 5 levels of validity, with level 5 representing the strongest (= causal) relationship and level 1, the weakest (= correlated) relationship. Depending on the goals of an outcome measurement, the practical possibilities and the availability of necessary resources, a decision must be made what level of 'validity' of the results is necessary or possible. It is not always possible (e.g. limitations in creating a design or availability of data) or necessary (e.g. management decisions on available resources) to establish a causal relationship (levels 3-5 MSMS).

The key to establishing a causal relationship is to find out what would have happened to the taxpayer in the absence of the tax administration's action; a *counterfactual*. So the tax administration needs a 'control group'. A control group is a group that has not received treatment and thus simulating the counterfactual. In this way the tax administration will know that the potential changes occurred due to its action and not due to other factors (outside the tax administration).

The most reliable way to carry out *formal experiments* (level 5 MSMS) and to get generalizable results is to draw a random sample to create a representative 'treated group' (or experimental group) and a representative 'control group' from the base population (see example 8)[18]. Randomization is also intuitively and methodologically a very straightforward way to identify causal relationships. These randomized control trials (RCT's) require putting thought into the design of the action beforehand, but there may not be a need for any extra efforts or campaign resources, as all one has to do is leave some randomly chosen individuals from the base population outside of the action reach.

Example 8: Level 5 MSMS, Randomized Control Trial

In 2017, the Polish Tax Office applied a randomized controlled trial (RCT) to test the effect of using letters to remind taxpayers to pay their taxes[19]. The trial had two objectives:

1. Test the effect of different behavioural messages on income tax compliance, and
2. Test whether the way the letters were delivered (regular versus registered mail) had an effect on compliance.

Taxpayers were randomly assigned to receive the official enforcement letter (a 'dunning letter') used by the Polish Tax Office or one of nine letters that were adapted using various behavioural designs. The findings were:

1. Behavioural letters significantly improved tax compliance relative to the dunning letter;
2. 'Hard-tone' messages were more effective than 'soft-tone' messages;
3. The effectiveness of some of the messages depended on taxpayers' characteristics;
4. Sending letters by regular mail (the cheaper option) proved to be just as effective as sending them via registered mail;
5. Tax compliance among taxpayers in arrears can be tackled cost-effectively.

While the validity of the results of outcome measurement benefits from randomly excluding some part of the taxpayer base from the treatment and examining that control group as the counterfactual, it is vital to make sure no statutory rights of taxpayers are violated.[20] For example, tax administrations often have legal requirements to treat everyone equally. Thus, it is good to ensure that excluding certain taxpayers from receiving information in the name of a randomised controlled trial is not an issue from a legal or ethical perspective.

If a formal experiment (randomization) is not possible, a *quasi-experiment* (levels 3 and 4 MSMS) can be set up in which random allocation is not explicitly done, but rather due to naturally arising factors in the setting. MSMS level 3 and 4 designs are thus called 'natural experiments'. The more we succeed in matching the treated group and the control group (making them comparable), the more the influence of external factors is diminished and the better differences between the groups can be addressed as due to the actions of the tax administration (and thus establishing a causal relationship).

Within the tax administration context, these naturally arising factors that generate quasi-random allocation are often found when considering the criteria that determine who are subject to the action. If the criterion is related to some 'continuous measure' of a taxpayer characteristic, such as firm revenue, regression discontinuity design[21] can

[18] Unwilling or Unable to Cheat? Evidence from a randomized Tax Audit Experiment in Denmark BEHAVIOR AL RESPONSES TO TAXPAYER AUDITS: EVIDENCE FROM RANDOM TAXPAYER INQUIRIES, National Tax Journal, March 2012, 65 (1), 33–58

[19] Applying behavioural insights to improve tax collection: Experimental evidence from Poland

[20] Tax Administration in OECD and Selected Non-OECD Countries: Comparative Information Series (2010), Compliance Risk Management Guide for Tax Administrations, Fiscalis Risk Management Platform Group, page 54

[21] Tax Compliance and Enforcement: New Research and its Policy Implications, University of Michigan, Field experiments in the developed world: an introduction, Oxford Review of Economic Policy, Volume 30, Number 4, 2014, pp. 585–596

be utilized to establish causal relationships (see example 9). Regression discontinuity design reposes on the basic idea that very close to the cut-off of the criterion, as in very close to the threshold of whether an individual is treated or not, individuals are similar enough to be compared.

Example 9: Level 4 MSMS, Regression Discontinuity Design (RD)

Since 2015, the Norwegian Tax Agency (NTA) has started risk-based audits in which individuals are selected for audits using the personal risk score of the year (based on a machine-learning model). For tax year 2014, NTA audited all taxpayers with a risk score above a certain threshold value. To measure the effect of this audit on future compliance, a Regression Discontinuity Design was performed by comparing the self-reported deductions of those who were just above the threshold with those who were just below the threshold. The assumption was that the two groups of individuals were similar and that they did not have the information of their risk score and thereby did not have the opportunity to manipulate the treatment status.
The results show that the audits have a strong effect on future compliance. The effect declines over time and becomes statistically non-significant in the third year after audit.

Sometimes there are *confounding factors*[22] that are associated with both the tax administrations action and the outcome a tax administration is interested in. For example, when looking at the effect of suing a firm for bankruptcy on the amount of tax debt collected, poor economic performance makes the firm likely to be sued for bankruptcy, but it also means that tax debts collected in the end are affected. In these cases level 4 can also be reached through *instrumental variables*.[23] An instrumental variable is connected to the action, but not the outcome. In the case of the example, an instrumental variable could be the random allocation of tax debt cases to more or less lenient clerks. The leniency of the clerk affects the probability of a firm being sued, but the leniency itself is not a predictor of the amount of tax debt that is ultimately collected. We can then randomize or observe random-like changes in the instrumental variable, which in turn causes a change in whether the tax administrations action is applied. This way the tax administration can isolate its action's contribution to the outcome, as it knows now that the change in the outcome was not muddled by the confounding factor, but due to the change in the instrumental variable.

If the criterion for exposing taxpayers to the action is not a continuous measure, but more a categorical indicator, such as the region, industry, or legal form of a firm, the tax administration can again compare one category of taxpayers that were subject to the action, to another category that was left out of the action reach, before and after the action took place. This method, called *differences-in-differences* (MSMS level 3), however, is not as robust in establishing causal relationships, as the tax administration has to make sure that the outcome of interest wasn't developing differently to begin with for taxpayers in the different categories (see example 10).

Example 10: Level 3 MSMS, Difference- in difference method

In 2009 the Italian Revenue Agency sent a letter to taxpayers suspected to have overreported some costs in order to decrease their taxable income without increasing the probability to be audited. A study based on a difference-in-difference method found strong and robust evidence that the letter did reduce manipulation of targeted costs, but that the impact on taxable income was much smaller.[24] This difference is likely due to a strategic response by the taxpayers who increased other costs that they perceived were not targeted by the Revenue Agency and were able to do so without increasing the probability to be audited. This shows that strategic responses are likely to appear also in an institutionally complex environment.

These identification strategies often come with more caveats and assumptions to fulfil than a randomized control trial, and the results are rarely as watertight. Therefore, it should be ensured that the setting meets the method criteria. Even though they might be methodologically more laborious, there is no need to be intimidated by these identification strategies. Sources of exogenous variation, leading to some individuals

[22] A confounding factor is a variable that influences both the dependent (= outcome) and the independent (= input or cause) variable.

[23] An Instrumental variable is a 'third' variable that is used when there are variables in the model that are influenced by other variables in the model.

[24] Evidence-Based Threat-of-Audit Letters: Do Taxpayers Respond Strategically in a Complex Environment?

being exposed to the treatment and some not, are often found in natural settings. Research designs for natural settings are also useful in measuring effects of actions that have already taken place.

If a research design aiming to establish causalities cannot be found or deployed, it is still possible to find information about *correlations* between the action of the tax administration and its results (MSMS levels 1 and 2). These correlation analyses include e.g. cross-sectional comparison of the treated and control group (see example 11), and before-after analysis of only the treated group. Correlation analysis is also useful with survey data, where novel information can be obtained about prevalent associations. The regression (between actions and results) can be tested by adding relevant controls (MSMS level 2)[25], but unobservable differences are likely to remain. Without answering the question what would have happened to the treatment group without the treatment no conclusions of cause and effect can be drawn.

Example 11: Level 2 MSMS, correlation analysis

In the study[26], a survey was sent to 250 randomly drawn individual taxpayers to examine the impact of a tax education program on tax payment compliance behavior in Lagos State, Nigeria. The survey was the Taxpayer Compliance Appraisal questionnaire. The questionnaire consisted of three parts: background questions, tax education program questions, and tax compliance behavior questions. The response rate was very high, 97%. The study finds that the tax education program focused on enlightening taxpayers on socio-economic implications of tax evasion opportunities and the transparent and accountable use of tax proceeds have significant influence on taxpayers' voluntary compliance.

Overview of pros and cons of the different designs

Table of different designs and respective strengths and weaknesses

Research design	Maryland scale	Strengths	Weaknesses
Randomized control trials	5	+ Offers the most credible way to create a counterfactual + Offers a reliable estimate of the causal effect + Results generalizable + Statistical analysis is simple	- Some extra effort and/or financial resources may be needed - Randomization may not always be possible due to e.g. legal restrictions
Instrumental variables	4	+ Can establish causalities with observational data + Helps address omitted variable and simultaneity bias	- Finding good instruments is difficult - Results may not be generalizable
Regression discontinuity	4	+ Can establish causalities with observational data + Settings are easy to find + Offers precise estimates (internal validity) + Ways to go around bunching issues	- Results may not be generalizable - Continuity assumption not testable

[25] A control variable is an element that is unchanged throughout an experiment and that allows the other variables being tested to be better understood.

[26] Taxpayers' education: A Key Strategy in Achieving Voluntary Compliance in Lagos State, Nigeria, Olowookere, J. K. & Fasina, H. T. (2013)

Research design	Maryland scale	Strengths	Weaknesses
Differences-in-differences	3	+ Settings are easy to find	- Parallel trends assumption is strong and not testable
Fixed/Random effects	3	+ Addresses unobserved heterogeneity	- Causal inference not conclusive - Requires panel/repeated cross sectional data
Matching	3	+ Can use observational data	- Causal inference not conclusive - Unobserved differences can persist - Requires a lot of data
Correlation analysis	2	+ Simple and cheap to estimate + Can still offer new insights (e.g. surveys) + Can be improved with controls	- Causal inference not conclusive
Before-after analysis	1	+ Simple and cheap to apply	- Causal inference not conclusive

TIP *If the outline for the planned action seems particularly tricky from a research design perspective, it is useful to refer to already existing analyses, such as those in the database of literature on tax compliance and risk management in this report. Existing results can also help determine correct sample sizes for e.g. randomized controlled trials.*[27]

TIP *Creating sample sizes large enough for obtaining statistically significant results necessitates the anticipation of the magnitude of the expected effect. If the tax administration has no prior experience in the action in question, it is useful to refer to other, similar experiments to determine what would be a realistic expectation of the change the action should induce.*

STEP 3 – STA EXAMPLE

STA's "Preventative Audits" program is based on random audits. The target population contains around 200 SMEs (sole proprietors or partners in trading companies) who filed late in 2009 within a certain region of Sweden. The treatment group is randomly selected and contains around 100 SMEs. The rest is used as a control group.

The research design for measuring the outcome in term of correct, complete and on-time filing is a 'Randomized Control Trial'. The effect is measured by comparing the average of outcome (i.e., declaration date, declared income, risk scores) of the treated group with the control group. The observed differences in average outcomes between the treated and the control group are interpreted as the effect of implementing "preventative audits".

A restriction of the research design is the small sample size that leads to large statistical uncertainty.

STEP 3 – IRA EXAMPLE

The design of the research had to deal with the risk based character of the audits as well as the several years processing time the audits took.

The targeted (treatment) groups were those taxpayers who had undergone different types of audits (soft audits or deep audits or sometimes both). The control group had not undergone any form of audit, but was selected in a way that made them similar to the targeted groups.

[27] Sample Size Calculations for Randomized Controlled Trials

IRA decided to choose a difference in differences approach (DID) in order to overcome the selection bias issues induced by non-exogenously assigned audits. According to the IRA auditing policies, taxpayers are selected for audits based on their compliance outcomes, which means that there are pre-existing differences between the audited and non-audited that can influence the outcome after audits. The DID approach allows accounting for those observable and unobservable differences between the two groups of taxpayers under the basic assumption that these differences do not vary over time.

Information on treatment and control groups was available before and after the audits occurred, in a panel data structure.

Restrictions of the research design:

- Some taxpayers could be audited in more than one year, which could drive them to behave differently from those who were audited only once. This could increase or decrease the impact determined by the first audit.
- The analysis did not include information about taxpayers audited before the risk-based audit campaign started. This could affect the control group and the results of the estimates.
- The analysis does not include the quality of the audit as perceived by the taxpayer.
- The analysis did not distinguish for any status of audit as ongoing or concluded.

3.3. Phase II: Implementation of the measured action

In this paragraph the steps to come to implementation of an outcome measurement plan are discussed. The implementation stage contains three important elements, ensuring data availability (step 4), securing the necessary resources (step 5) and starting with the actual intervention/action (step 6).

3.3.1. Step 4: Ensuring data availability

In this step, at the latest, it is time to make sure all necessary data are within reach. Tax administrations often collect and store many kinds of internal data related to their own operations[28] such as tax return and payment data. It is important to be clear about what data or datasets are needed in what format and where/when to get them within the organisation.

If there is a need for external data, it is essential to know the local data access rights of a tax administration including privacy aspects.[29] The relevant data source depends on the specific project and the industry in question. Banks and credit card companies, for example, can have useful comparative data. It is important to allow enough time for the external parties to collect and deliver the requested data, as well as to validate the data upon receipt.

A third possibility is to get data through issuing surveys (and/or interviews). Tailored surveys allow pinpointing the exact questions the tax administration wants to be answered, and obtaining novel information that is not available otherwise. The reliability of the survey data can be improved by using an external agency as an intermediary to collect data on behalf of the tax administration. In such a way the anonymity of taxpayers can be guaranteed. Surveys can be used to clarify the preferences of taxpayers, the attitude towards taxes and tax evasion/tax avoidance and voluntary compliance.

[28] Compliance Risk Management Guide for Tax Administrations, European Commission (2010), Fiscalis Risk Management Platform Group, page 32

[29] Tax Administration in OECD and Selected Non-OECD Countries: Comparative Information Series (2010)

Points to consider are:
- potential changes in reporting standards or data collection practices that may result in imperfect or incomparable data;
- whether a database is versioned or only saves e.g. the latest version (overwriting previous versions). In this case, it is even more crucial to plan the data resources carefully before any actions are taken, so the values prior to the action can be captured;
- possible legal constrictions (to use or combine data).

TIP Data scraping and web crawling are also available nowadays to collect data that might have been laborious to obtain a few years ago. Before putting these tools to use, it is good to ensure that scraping is permitted from the desired source and that relevant data protection legislation is adhered to.

STEP 4 – STA EXAMPLE
All necessary data needed for this study were available in STA's internal database. To begin with, the analyst identified the target population by using the "Base Register" for firm type and geographic region and the "Taxation register" for the declaration date in 2009. Thereafter, a randomized selection of the treatment group was performed.

For each of the SMEs in the target population, the following data for tax year 2009 and 2010 was collected from STA's Taxation register:
- Declaration date
- Declared taxable income
- Declared turnover and cost

An audit memorandum was prepared for the tax auditors to fill in relevant information collected during the audits, regarding e.g. the mistakes discovered.

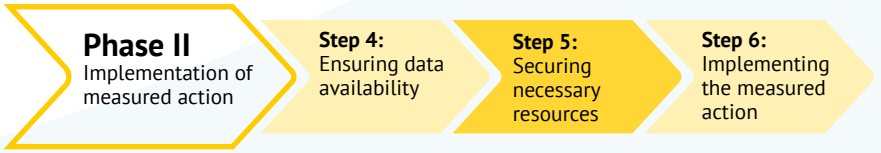
STEP 4 IRA EXAMPLE
The dataset was drawn from two internal sources:
1. Tax audit records
2. Tax returns register

From the Tax returns register, all information on Regional Tax on business activities , VAT and PIT was included in the research. From the Tax audit records, all information from the audit records was also included in the research.

The research was conducted on a 10% sample of the reference population of self-employed and small businesses (about 750.000 taxpayers), audited and non-audited taxpayers included.

The computer scientist verified the quality of the data and its format being suitable for subsequent exploration. During this verification process simple mistakes, outliers and incoherencies were discovered.

3.3.2. Step 5: Securing necessary resources



After having formed an outcome measurement plan, the practicalities of incorporating outcome measurement into the projected action need to be considered. This means identifying the necessary resources to carry out outcome measurement successfully, including a realistic timeframe for the execution of the measurement.

In terms of required time, not only how long the action will take, but also how long it will take for the outcome to become observable is important. In addition it is also important to reserve time for data analysis. This means having the right experts available to cover the measurement process. Data analysts, statisticians and economists often have the deepest understanding of the methodology, but understanding the context of the measurement and the policy implications of the results is equally important. This is why a comprehensive outcome measurement process also necessitates policy, business, and legal experts. Naturally, the implementation of the action itself also needs operational staff, which should be taken into account, for example when considering how many audits would have to take place for a sample size to be large enough.

TIP *If there is a need for more advanced econometrical analysis or modelling beyond the tax administration's resources, researchers specialized in tax issues at universities and research institutes may be willing to help, if in exchange they can utilize the work conducted in a research paper.*

STEP 5 – STA EXAMPLE

Staff resources

For implementing the preventative audits, five tax auditors were needed, and working hours were planned to be 500 hours in total.

For the outcome measurement of STA's preventative audits, two type of resources were needed: human resources and equipment. The team involved in the analysis included a coordinator and two analysts with different expertise. The coordinator worked with project management and served as a liaison between the tax auditor and the analysts. The analysts had relevant knowledge in statistics, econometrics and policy evaluation. They were also familiar with the STA's database and were able to work with SQL to prepare the dataset for the analysis.

The equipment required in term of software refers to: SAS, Excel, R (open source free to use).

Timeframe

January 2010 – Identify the target population and select the treatment group
February – April 2010 – Perform the preventative audits
August – December 2010 – Perform outcome measurement
January 2011 – Deliver the report

STEP 5 – IRA EXAMPLE

Staff resources

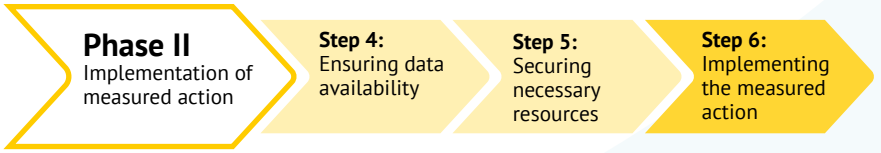
IRA outcome evaluation encompassed two types of resources: human resources and equipment. The full team involved in the analysis phase included: a computer scientist and three economic analysts with different expertise. The computer scientist was necessary for data preparation to make the dataset for the subsequent analysis available. The economic experts were acquired for their skills in Applied Econometrics and policy evaluation, one tax auditor and one tax expert for their business knowledge.

The software used for running the outcome evaluation consisted of software for econometric and other statistical analysis: SPSS, Stata, R (open source free to use).

Timeframe

It took approximately two years from the start of the evaluation to the final report.

3.3.3. Step 6: Implementing the measured action



In this step the actual action or treatment the tax administration wants to carry out and whose effect is being measured, is implemented.

For the analysts and other experts who are carrying out the outcome measurement in practice, the role in this step is to make sure everything is falling into place correctly as planned in the previous steps. While in this step the operational staff is in the lead (carrying out the planned action), it is important for the analysts to communicate with the experts implementing the action to be aware of e.g. any glitches that should be accounted for in the data analysis.

During the implementation, it is recommended at this point to pay attention to appropriate data entry and storage (see step 4), and to monitor the adherence rate through keeping track of individuals dropping out of the treatment.[30] For example, if a tax administration plans to sample out 500 firms for audit, and audits them all, the adherence rate among the group that was sampled for the 'auditing treatment' is 1. However, if a tax administration sends out 50 000 letters informing the taxpayers about a new and efficient way of filing taxes, it is not certain how many of the taxpayers opened and read the letter, thus receiving the 'information treatment'. If, in the end, 10 000 taxpayers used the new and efficient way of filing taxes, the tax administration might conclude that the ratio of taxpayers adopting the new filing channel to the informed taxpayers was 1:5. However, if only 20 000 taxpayers opened and read the letter, the information treatment, in fact, yields a ratio of 1:2. If it is not possible to observe the adherence rate, the estimate, in fact, describes the *intention-to-treat*[31] *effect*, rather than the average treatment effect.

STEP 6 – STA EXAMPLE

During the implementation of the preventative audits, the auditors were instructed carefully by the program coordinator that they had to report the process and findings

[30] The Econometrics of Randomized Experiments, chapter 9
[31] Intention to treat (ITT)

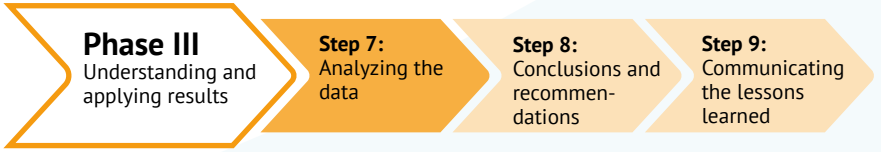
of the preventative audits and note any dropouts, and that they had to follow the same audit routine, as well as fill in the audit memorandum properly. In this research, 10 dropouts were noted during the implementation. Reasons for dropouts were also noted for later analysis in order to check whether validity of this randomized control trial is threatened as those who completed the process may differ from dropouts in a systematic way.

STEP 6 – IRA EXAMPLE
IRA's risk-based audit campaign has been conducted on a risk-based selection of self-employed and small businesses. Either the local or the central Office was in charge of making decisions about numbers and types of audits. Auditors made their investigations over their jurisdictional area. All auditing results were reported in a central tax audit register. After the campaign a further quality check of available data was conducted to verify outliers and consistency of all information acquired.

3.4. Phase III: Understanding and applying the results

In this paragraph the steps to analysing and communicating the results are discussed. The results stage contains three important elements, analysing the data (step 7), conclusions and recommendations (step 8) and communication (step 9).

3.4.1. Step 7: Analysing the data



After the measured action has taken place and all raw data has been collected, data management must be carried out to ensure the data are ready for analysis.

With the raw data in place, one can now provide an analysis on the data quality and produce a data quality report. The report can describe how many observations were deleted and why, how many values were substituted, why and how, and if the number of observations is sufficient for the analysis.

After the data are cleaned they can be analysed. Thanks to the careful planning and the work put into the previous steps, the analysis phase is putting together all the pieces laid out in the previous steps using statistical methods. In this stage is to be confirmed that the methods and applied software suit the selected design.

Analysis of data starts with descriptive statistics illustrating the basic characteristics of the data, e.g. the distribution, summary statistics, outliers, trends etc. Next – based on the research design that was chosen in step 3 – the underlying model is fitted with the data, using statistical techniques (e.g. regression, comparison of means). The results of the model can be tested for statistical significance and assumptions and robustness (depending on the research design). When this work is done, the findings of the data analysis are summarized and converted into a user-friendly format.

It is important to be transparent about the limitations in the analysis, as it determines the confidence in the results. Were the assumptions of the used methodologies met? For example, if data was collected through a survey, attention should be paid to the response rate.

External validity concerns should be considered as well. How generalizable are the results? For example, if an auditing campaign has only looked to a certain industry, the results will reflect the behaviour of taxpayers in that industry but not be representative for the entire economy.

There could also be adverse or unexpected effects in other indicators in addition to the primary object of the project. As these possible 'side effects' were already identified in the effect chain in step 2 and the data obtained for in step 4, verifying these side effects should not be a problem.

In addition to the estimation of the causal (MSMS 3-5) or plausible (MSMS 1-2) effect of the treatment, it is also informative to include a cost-benefit analysis of the action *(see effect chain)*. As resources are scarce, the cost-benefit analysis often proves to be a valuable part of the analysis from the operative perspective. For the cost-benefit analysis, direct costs as well as opportunity costs should be included (example 12).

Example 12: Cost-benefit analysis

A tax administration has conducted a letter campaign encouraging taxpayers to file their tax returns online instead of on paper forms. From a direct-benefit perspective, there is most likely no significant increase in tax revenue. However, the opportunity costs of tax officers time must also be considered: when tax officers are not handling paper returns, their time is free to focus on other tasks (e.g. tax audits).

STEP 7 – STA EXAMPLE
The analysis started with a description of the audit results by examining the audit memorandum. It appeared that the majority of the SMEs made mistakes in their declarations, for example, putting private expenses in the business accounting records and/or having not filled the declaration form completely. However, the tax administration didn't make additional assessments since the audits were made before filing. The SMEs instead were given an opportunity to correct the mistake(s) themselves.

Next, the analysts measured the outcome of the preventative audits in term of improvement in filing on time, correct filing and increase in tax paid. Thanks to the design of the randomized control trial, the effect of the 'preventative audits' could be measured by comparing the outcome of the treated group with the control group. Statistical tests were also performed to verify whether the effects found were statistically significant. It appeared that preventative auditing has a positive effect on correct filing. More than half of the audited firms that made mistakes in their declaration forms made the necessary corrections. More importantly, the treated group on average was less likely to be selected for "Risk-based audits" compared to the control group after treatment.

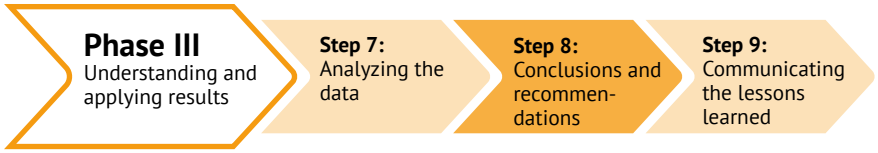
However, no statistically significant effect was found for filing on-time and increase in tax paid. This may be due to the fact that the sample size was too small to make reliable statistical inferences.

STEP 7 – IRA EXAMPLE
Descriptive statistics were drawn to get a feel for the business characteristics in order to better understand the treatment group and the control group. This was helpful to better interpret the outcome measurement and the possible differences between the treatment and control group.

Next, the analysts carried out the actual evaluation by applying econometric techniques suitable for a proper evaluation estimation. According to the data availability, a difference in differences approach was applied. In particular five outcome variables were examined to understand taxpayers' behaviour with respect to various taxes: IRAP base and tax declared, VAT turnover, VAT declared, PIT declared. IRA compared the trend of these variables in the treatment group with the control group. The assumption was that the trend in both groups, before audit, would be the same so differences occurring after audits could only be caused by the audits. This crucial assumption for the validity of the estimates is called *common trend assumption*[32] i.e. the condition for the validity of the estimates requires that the pre-audit trend of the audited taxpayers is the same as the trend of the non-audited taxpayers group.

[32] D'Agosto, E., Manzo, M., Pisani, S., & D'Arcangelo, F. M. (2018). The Effect of Audit Activity on Tax Declaration: Evidence on Small Businesses in Italy. Public Finance Review, 46(1), 29–57

3.4.3.Step 8: Conclusions and recommendations



Outcome measurement should aim to produce high-quality information that can be used in improving the tax administration and helping in real-life decision-making. Thus, measurement results should be translated into practical conclusions, keeping in mind the robustness of the results, established in the previous step.

In this step the results of the outcome measurement are discussed and recommendations for applying the results in practice are made. Internally, managers, policymakers and business experts must be involved in discussing the results to close the CRM circle and to safeguard a justified and fruitful use of the results. Aspects that need to be considered are:

- What kind of impact can be attributed to the action based on the results?
- Have the objectives been reached?
- What is the validity or robustness of the results?
- What are the policy implications according to the analysis?
- What are the limitations that should be considered?
- Should the tax administration continue with the policy or not?
- Are there still some aspects that need to be measured, or re-measured, to reach conclusive and sustainable results?

A resulting policy recommendation may be made. Recommendations might include the establishment of a new instrument, discontinuing an existing policy, changing an internal process, adjusting modes of external communication, amending an existing law or maybe receiving new legal rights to access third party data. The recommended next step can also be to conduct more measurement to obtain results that are more precise.

In general, the stronger the results are in terms of statistical validity, the more confident you can be with making policy recommendations. Less decisive measurements or circumstantial evidence call for reasoning about the degree of certainty of the results. A suggestion is to include all relevant information related to the interpretation of the results when forming practical conclusions or making recommendations. Such information can include the composition of the target group, the degree of certainty that the results are not capturing other factors outside the tax administrations action of interest, and any limitations in research design, data or the analysis conducted.

The results of the outcome measurement could disclaim the hypothesis, or reveal that the action yielded no effect or even adverse effects. In this sense, a statistically significant result of 0,00, is very valuable, as it can be a trigger to reassign the resources the ineffective action is using. It can take courage to highlight and follow the measurement results even when it means discontinuing something; but it should still be remembered that the power of outcome measurement – and data analysis in particular – lies in revealing and producing information a tax administration wouldn't have had otherwise, or would have to rely on potentially inaccurate assumptions and gut feelings. If more evidence or results of higher precision and validity are needed, it is always possible to repeat the experiment on a slightly different demographic or a larger sample, whatever the concern might be. In any case, a practical application that is implemented based on the information received through outcome measurement, should also be evaluated to make sure the desired outcomes are still being reached.

STEP 8 – STA EXAMPLE

Since this outcome measurement showed that preventative audits have a positive effect on inducing correct filing, these findings were used as an empirical support for STA to develop two new instruments:

- Control of documentation obligation. Subsequent legislative changes gives the Tax Agency the right to perform a similar examination as 'preventative audit' without the need of an 'Auditing Warrant'. This type of control does not require to be carried out by an auditor.
- Control of newly started businesses as well as limited companies who opt out of hiring accounting firms to prepare their tax declarations.

By using these two new preventative instruments, STA expects to induce correct filing and thereby improve voluntary compliance with relatively lower costs than performing repressive audits after filing.

STEP 8 – IRA EXAMPLE

IRA findings suggest an increase in the compliance proxies (tax base and tax declared) used. The magnitude of the effect differs across taxes; an audit increases the regional

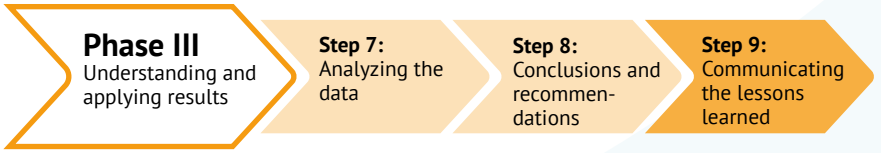
business tax declared by 1.8%, the VAT by 5.3%, and the PIT by 5.2%. Likewise, the persistence of the effect of the audit over time also differs across types of tax. While for both regional business tax and PIT the effect remains positive also in the first year after the audit, in the case of VAT this effect seems to vanish in the first year following the audit.

Deep audits turn out to be the most effective action to increase tax compliance. In particular, taxpayers who experienced a deep audit increased the regional business tax they declared by about 19% in the year of the audit, the VAT declared by about 18% and the PIT by 14.7%. Soft audits seem to exert a weaker effect on tax compliance. However, as they often serve as input to subsequent deep audits, their effect may be underestimated.

Based on these results of the research, the administration was recommended to continue with the annual campaign.
Further on the findings were made available for:
- discussion with offices involved in taxpayers' selection rules;
- the next audit campaigns measurements;
- the audit campaigns measurements at different territorial level with a view to improving the knowledge about audits by Local offices.

It was also recommended to compare these results with those from others instruments, i.e. letters or other preventative instruments.

3.4.4. Step 9: Communicating the lessons learned



Communication of the results (conclusions and recommendations) multiplies their benefits. Both internal and external stakeholders could benefit from communicating results and lessons learned. In general, for successful communication one could consider e.g.:

- Who are the stakeholders?
- What is the (differentiated) message we want to bring across (to various stakeholders)?
- What do we want to achieve with the communication (e.g. additional resources for future action, a law amendment, influencing taxpayer behaviour)?
- What can or needs to be said (internal and external) about the research plan, the process and the results to whom, and what information is confidential?
- What are the lessons learned?
- What is the best forum or channel to get through to the targeted group(s)?

A tax administration could use the results of outcome measurement in various ways. Using the results the administration may learn and improve its own compliance risk management strategy, manage staff by making more educated choices with regard to capacity, show the added value of its activities (accountability) and reflect on the feasibility of legislation with the legislator or politicians (see paragraph 2.2.3).

Informing stakeholders on the performance of the tax administration is a responsibility of the administration. However, communicating to the tax administration's stakeholders in an annual report, for example, can also increase positive attitudes of taxpayers towards the tax administration, as it is beneficial to show that the tax administration is reviewing its own actions, adjusts its strategy when necessary, and is mature enough to act on the results. This can be seen as a token of responsibility and respect for the taxpayers' money. It can enhance commitment to the CRM strategy as well, and change the attitude towards more modern insights on compliance management. An annual report can be the place

to address legal matters as outcome of the reflective function of outcome measurement.[33]

Other teams, units and departments, as well as tax administrations in other countries could benefit when they know of the new insights and potentially find new applications for the results.

In general, however, results of projects and trials should be communicated afterwards, so potential treatment groups are not affected with premature information Simply being aware of being observed might alter the behaviour of the individuals in the group of interest. As this *Hawthorne effect* is difficult to quantify, it is best to limit the treatment group's knowledge of being observed during the trial, where possible.

In addition, when communicating statistical parts of the analysis, attention should be paid to using clear language and highlighting the main results while also keeping in mind the caveats. Special attention should be directed to the nature of the connection of the factors under analysis. Are the results generalizable? Has a causal relationship been identified in a valid way? If not, the communicative efforts should abstain from the use of expressions such as 'effect' and 'cause', and rather use expressions related to 'correlation', 'connection', and 'association'.

A tax administration should aim to be clear and easily understood by its target group, and it's important to be transparent about the experiment as well: give justification concerning the choices that were made during the measurement planning phase, explain who were included in the experiment and who were left out, and what implications this has, and whether there are other caveats that should be considered. The communicated message shouldn't be just figures and graphs, it should be about creating an easily understood narrative, and how the measurement process has yielded valuable information.

Communicative efforts also include adequate documentation of every step of the process. This way knowledge can accumulate within the organization, and improvements and future endeavours can build on existing experiences, rather than having to go through the same challenges multiple times.

STEP 9 – STA EXAMPLE
The findings of the experiment were first communicated to the head of the auditing function.

Then the results and the practical conclusions and suggestions for next steps and how to utilize the results through new instruments were presented to the management of the tax administration.

The evidence gathered in the experiment was also presented to the policymakers to support a legislative change towards less rigid modes of audit.

STEP 9 – IRA EXAMPLE
The results were presented to the top management, including the Directors of the IRA Departments, to the audit operational unit within the IRA and were published in the Public Finance Review.

[33] The relationship of the tax administration with lawmakers differs so the channel to put legal issues on the agenda can be more or less direct.

3.5. References

If you want to know more, you might find these resources helpful:

- Compliance Risk Management Guide for Tax Administrations, European Commission, Fiscalis Risk Management Platform Group, 2010
- Measures of Tax Compliance Outcomes: A Practical Guide, OECD Publishing, 2014
- Tax Administration in OECD and Selected Non-OECD Countries: Comparative Information Series (2010)
- Guide to scoring methods using the Maryland Scientific Methods Scale, What Works, Centre for Local Economic Growth
- Mastering 'metrics: the path from cause to effect, Angrist, J. D., & Pischke, J.-S., 2015
- The Econometrics of Randomized Experiments, Cornell University, Susan Athey, Guido Imbens, 2016
- Sample Size Calculations for Randomized Controlled Trials, Johns Hopkins Bloomberg School of Public Health, Janet Wittes, 2002
- Field experiments in the developed world: an introduction, Oxford Review of Economic Policy, Volume 30, Number 4, 2014, pp. 585–596
- Tax Compliance and Enforcement: New Research and its Policy Implications, University of Michigan, Joel Slemrod, 2015
- Evidence-Based Threat-of-Audit Letters: Do Taxpayers Respond Strategically in a Complex Environment?, Carlo, Fiorio & Alessandro, Santoro, University of Milano-Bicocca, Department of Economics, revised 26 Sep 2017
- Behavioral responses to taxpayer audits: evidence from random taxpayer inquiries, National Tax Journal, March 2012, 65 (1), 33–58
- Unwilling or Unable to Cheat? Evidence from a randomized Tax Audit Experiment in Denmark, 2009
- Taxpayers' education: A Key Strategy in Achieving Voluntary Compliance in Lagos State, Nigeria, Olowookere, J. K. & Fasina, H. T. (2013)
- Intention to treat (ITT), European Patients' Academy, 2015

4 Examples and literature

The fourth section, Examples and literature, provides a link towards a large collection of practical examples and academic literature in the field of outcome measurement, combined in a 'Database'.

This section – after a brief introduction in par. 4.1 - in par. 4.2. explains how to access this database and how to make adequate selections. Furthermore, a predefined selection, based on two criteria – Level of management & Type of measurement –provide the reader quick access on finding useful academic studies.

4.1. Introduction

The section, Examples and literature, gives an overview of practical examples of using outcome measurement within the EU tax administrations. In addition, a large volume of academic literature has been made available. The academic literature can be used in various ways: it serves as 'evidence' for assumptions tax administrations are making in their Compliance Risk Management process, it also can be used as an example of how to carry out outcome measurements and to gain knowledge on academic research in this field of expertise.

The database is not limited to the documents that were referred to in the previous sections of this report but also contains a large body of other literature, covering both measurement of proactive and reactive actions of tax administrations at different levels of management. However, this list is not exhaustive. Through the *link to this database*, the reader does not have to leave this document to have access to all background information. In the interactive document the reader can easily move from the previous chapters to the examples and literature in this chapter and back.

4.2. The Database

The database is presented in the form of an Excel workbook. It has four main components:

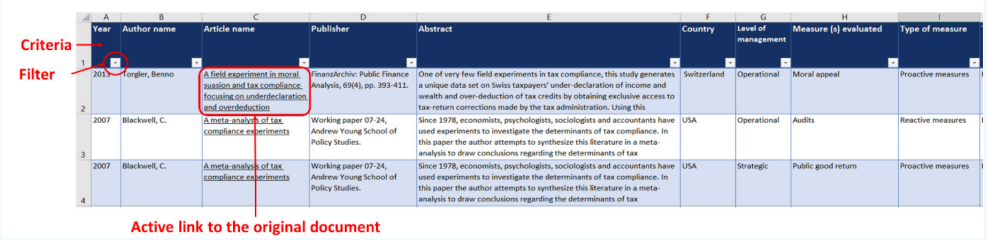
1. Academic literature in which compliance outcomes of different instruments are measured empirically by researchers and published as articles, chapters in books or working papers.
2. Practical examples of outcome measurements carried out by tax administrations. These examples were obtained by our workgroup after having contacted these tax administrations. In general, these examples have not been externally published. We provide therefore a summary of the main content.
3. EU Commission or OECD Reports in the field of Compliance Risk Management and Outcome measurement.
4. Methodology literature that are useful to understand the statistical methods or research designs for measuring outcome.

4.2.1. Description of the database

In the Excel workbook, these four main components are presented in four separate worksheets named as follows:

- Academic literature
- Practical examples by tax admin
- EU, OECD and other documents
- Methodologies

The literature pieces in each of these sheets are stored as records in the rows. The relevant and important information about each literature piece is presented in the columns. The information of the columns to search the database will be referred to as 'criteria'. Each column is fitted with a filter that can be used to decide which rows in the worksheet to display. These filters can use data such as conditions or values, as in any Excel worksheet.



4.2.2. Content of the database

Common columns for all the sheets include the basic information such as:

- article name (with a **link to the original document** when available),
- author or institution name,
- year of publication and
- abstract.

For the effect studies of academia and the practical examples of tax administrations, we also include some attributes that give information about the *instruments* and the *methods* that were used, and about the *findings*.

Columns with information about the **instruments** measured are:

- *Measure evaluated*: a keyword describing the nature of the instrument (see also the table at the end of this section for the list of all instruments mentioned in the sheets *Academic literature & Practical examples by tax administrations);*
- *Country*: the country in which the instrument was implemented;
- *Level of management*: whether the instrument is a strategic, tactical or operational one (see also at the *quick access* for an explanation of this criterium);
- *Type of measure*: whether the instrument is proactive or reactive (see also at the *quick access* for an explanation of this criterium).

Columns with information about the **method** used to perform the measurements are:

- *Type of data*: experimental, administrative or survey data;
- Type of experiment: field experiment, natural experiment, natural field experiment or quasi-experiment;
- *Research methods*: what type of research design is used to identify the causal relationship between the instrument and the outcome, for example, RCT, matching, difference-in-difference, and panel data with fixed effect and so on;
- *Maryland Scale*: the level of Maryland Scale the research methods relied on;
- *Outcome variable(s)*: the variable (indicator) used to reflect the change in compliance behaviour, for example, filed income, payment rate etc.

Columns with information about the **findings** include:

- *Direct effect*: whether the instrument proved to be effective on improving compliance for those treated;
- *Spill over effect*: whether the effect on improving compliance is found among those who were not directly treated but in one way or another related to those treated.

4.3. Quick access

In addition to the four sheets described in par. 4.2 , the database has also a coversheet named Overview. This overview table, also represented at the end of this section, gives an overview of all the **instruments** mentioned in the sheets Academic literature & Practical examples by tax administrations, organized around 2 criteria: 'Type of instrument' and 'Level of management'.

1. Type of instrument: Proactive versus reactive actions:
   - By ***proactive*** we mean: all actions intended to prevent non-compliant behaviour ***before the tax return has been filed***;
   - By reactive we mean: all actions deployed ***after the tax return has been filed*** and non-compliant behaviour has been detected.

2. Level of management: these levels are[34]:
   - ***Strategic level***: on this level, the most important external and internal risks facing the administration are identified and forms the basis for strategic decisions made by the management with a 3 – 5 years horizon; we're talking about taxpayers in general; for example: the administration wants to invest in the 'education' of taxpayers.
   - ***Tactical level***: on this level, the strategic risks are broken down in specific risk areas which should be treated on an annual basis; we're looking here to a specific segment of taxpayers; for example: a specific documentation, explaining tax obligations, will be provided to starting business.
   - ***Operational level***: on this level we encounter the actions concerning individual taxpayers; for example: a reminder to starting businesses failing to file their tax return on time.

	Strategic	Number of documents: Academic literature	Number of documents: Practical examples	Tactical	Number of documents: Academic literature	Number of documents: Practical examples	Operational	Number of documents: Academic literature	Number of documents: Practical examples
Proactive	Education	3		Cooperative compliance		1	Accounting controls for new firms		1
	Information of public expenditure	3		Filing aid		1	Letter Campaign		2
	Pre-populated tax returns	1		Information of audit rule	4		Information of increased enforceme	1	
	Public good return	2		Media Campaigns	1		Information of penalty	2	
	Reform of tax administration	2		Model for itermediairies		3	Information of public expenditure	1	
	Simplification of tax code	2		Model for new businesses		1	Information of social norm	6	
	Split payment of VAT	1		Pre-populated tax returns		1	In-person visit	1	
	Tax code provision	1		Supervision	1		Moral appeal	12	
	Tax rates	5		Third-party information	1		Preventative audits		1
	Third-party information	2					Rewards	1	
							Soft reminder	3	
							Supervising and visiting new firms		1
							Text message campaign		3
							Threat to audit/enforcement	10	
Reactive	Penalty	6		Risk-based audit programme	1		Audits	22	2
							Criminal investigation	1	
							Letter Campaign		1
							Personal contact with tax payers th		1

⇒Access the entire database of academic literature

⇒Access the entire database of practical examples by tax administrations

⇒Access the database "Methodologies"

⇒Access the database "OECD documents"

Figure L - Overview table of the documented instruments in the database.

Figure L *- Overview table of the documented instruments in the database.*

By clicking an instrument in this overview table, you will be directed to the literature for this specific instrument. Clicking on the "⇒Access…" entries will direct you to the entire list for that category permitting you to make your own selection.

Please notice that for most items in the database there is an embedded link that will direct you to the original document. If you wish to read the full text, just single-click the name of the article.

Clicking the icon Overview on any sheet, will redirected you to the overview table.

You can get access to the database by double-clicking this link:

https://webgate.ec.europa.eu/pics/filedepot/22464#

[34] Fiscalis Risk Management Platform: EU Compliance Risk Management Guide for Tax Administrations (2010), section 5.8.

5 Executive summary and suggestions

The fifth section, Executive summary and suggestions provides a management summary in which the most important content is highlighted and some suggestions for a successful implementation of "outcome measurement" within a Compliance Risk Management strategy are proposed.

5.1. Executive summary

In this paragraph the main elements of outcome measurement are summarized.

- A continuous improvement of the CRM process
  Outcome measurement is part of the CRM process who's own objective is to enable tax administrations to accomplish their strategic objectives by facilitating management to make better decisions. Outcome measurement is a form of evaluation that describes if and to what extent the organisation's objectives on different levels are achieved. Therefore, applying outcome measurement systematically will continuously improve the CRM process as a whole and increase the chances of achieving these objectives.

- Clear objective setting
  SMART objectives on strategic, tactical and operational level are a prerequisite for outcome measurement. Carrying out outcome measurement has a disciplining effect because it forces the tax administrations to be more clear on the goals they want to achieve (par. 2.2.3.). This will facilitate the communication about the strategy and the goals, both internally and externally.

- Think big, start small
  Changing society forces tax administrations to adjust their CRM and to meet the demand for more accountability (see par. 1.1). Outcome measurement is part of the answer to these high stakes, but it is also a process that should be well planned (see chapter 3). However, outcome measurement can be started small with just one action, it is not necessary to try to measure all actions at once.

  Neither is it necessary to aim immediately for the perfection (established causal effect by randomized control trial). A correct established correlation can be a valuable first step in the process (see 9 steps methodology explained in chapter 3).

- Supports management in the making of the right choices
  Outcome measurement shows which activities turn out to be more or less effective. It provides management with evidence-based information so that it can adjust its activities in the direction of the most effective ones. In other words, outcome measurement enables management to make educated choices with regard to capacity and the most effective instruments.

- Story telling
  Communicating the evidence-based results of the outcome measurement in a clear narrative. The narrative has to describe the broad results of an activity, a project or group of taxpayers, instead of just focusing on one or two KPI's: what kind of activities did the administration undertake, what were the resources used, what was the output, what were the side-effects of the activity, what external factors were of influence, what changes in the behaviour of taxpayers were observed (outcome)?

  It presents the tax administration as modern, state of the art, logical and critical towards itself. It induces trust and credibility towards society and shows the tax administration's added value. Also, in doing so, the story telling enhances belief in the applied CRM strategy and bolsters self-esteem.

- Sustainable impact on behaviour (or the outcome of outcome measurement)

  Starting with outcome measurement, one must be aware that expected results cannot always be realised on the short term. To see long term effects, outcome measurements need to be repeated.
  The objectives trees and effect chains, described in the methodology (steps 1 and 2), draw attention to this issue. They will identify both long-term, medium- and

short-term outcomes as well as give information on inputs and outputs. They will also help to break down larger and more complex problems into more manageable pieces which can be considered to be a short term results (quick-wins).

5.2. Suggestions

In this paragraph some recommendations for outcome measurement implementation are made, based upon the challenges tax administrations face in practice.

- ### Developing a vision on outcome measurement
  If the tax administration plans to develop outcome measurement, it is suggested to state this clearly in its strategy and to integrate outcome measurement in its CRM process. The development then becomes a part of the day-to-day activity that can be followed systematically and guarantees that outcome measurement will not be a stand-alone activity.

- ### Management commitment
  The support of the management regarding to the implementation of outcome measurement is the primordial prerequisite. This means the management agrees to assign the often limited necessary human resources to this process.

  Secondly, management must participate actively in phase 1 of the methodology, that is, setting the objective(s) and deciding on the desired level of precision of the results.

  In the end, the management has to commit to act upon the results of the outcome measurement and to use the right tone of voice (not "right" versus "wrong").

- ### Cultural aspects
  In practice, a tax administration may often rely on potentially inaccurate assumptions and 'gut feelings'. The power of outcome measurement lies in revealing and producing evidence based information a tax administration wouldn't have had otherwise. If the results of outcome measurement reveal that the action yielded no effect or even adverse effects, this is equally important information and should be a trigger to reassign resources and be more efficient. It is therefore suggested not to withhold this information but to be transparent about it.

- ### Stakeholders
  Communication of the results multiplies their benefits. It is suggested to communicate to both internal and externals stakeholders for example through an annual report. This can increase positive attitudes of taxpayers towards the tax administration, as it shows that the tax administration is reviewing its own actions and adjusting its strategy when necessary and is mature enough to act on the results. This can be seen as accountability and self-reflection.

- ### Application
  A tax administration can deploy different types of actions. It is important to apply a measurement outcome on all types of actions (old and new, preventive and repressive). Especially when a tax administration starts with a new (preventative) action, it is recommended to start with outcome measurement on this action, as a new strategy is often challenged by stakeholders and needs proof.

6 Annexes

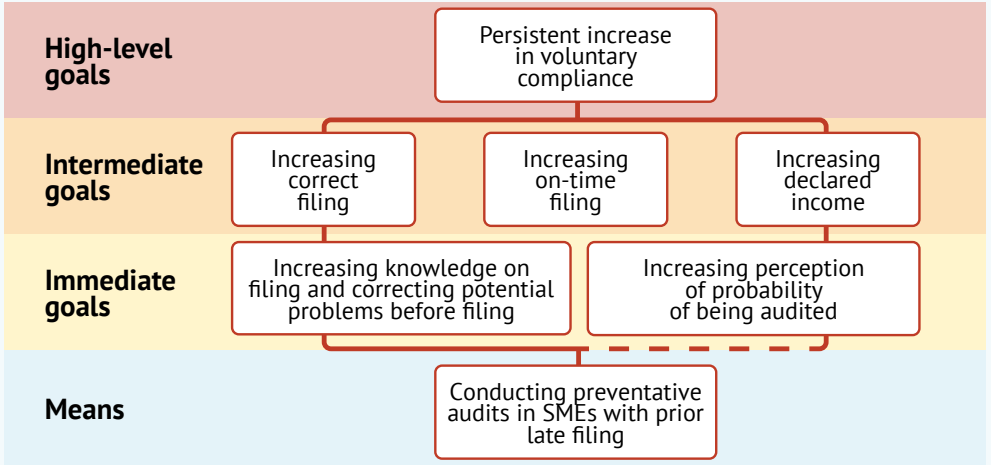
6.1. STA Example

STEP 1 – Objective setting[35]
In 2010, the Swedish Tax Agency launched a program called "Preventative Audits". In this program, on-site audits were carried out prior to the filing due date for Small and Medium-sized enterprises (SME), within a certain region, who filed late in 2009.

The program is a one-time trial and the purpose is to understand whether it is effective to promote "voluntary compliance" with this new way of working.

The objectives that STA wanted to evaluate are whether preventative audits can lead to:
- Increased correct filing
- Increased filing on-time
- Increased declared incomes/taxes

On a more strategic level, these objectives contribute to an increase in voluntary compliance, as seen in the objective tree :

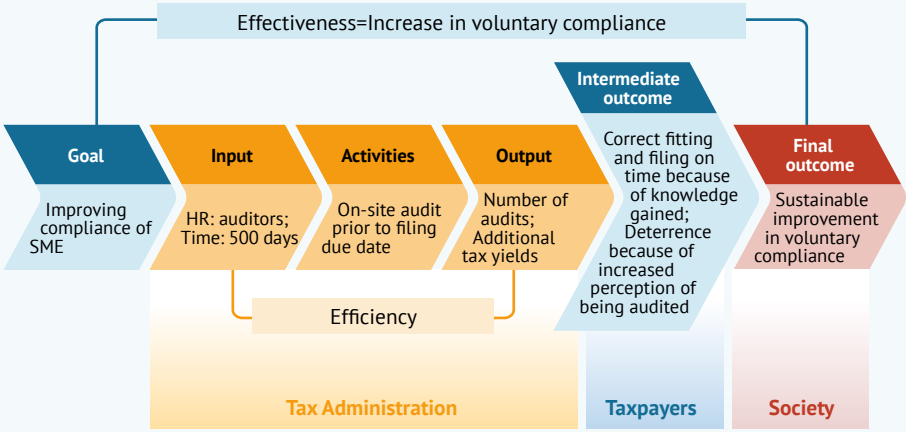


[35] The running example is based upon a practical example of the Swedish Tax Agency (STA), which is modified for educational purposes.

STEP 2 – Forming a hypothesis
STA's "Preventative Audits" were expected to have two main effects on tax compliance: a direct effect and an indirect effect. The direct effect comes from the mistakes directly detected and prevented during audits before the filing due date, therefore resulting in additional collected revenue. The indirect effect is related to changes in the behaviour of the audited taxpayer after the auditing and can be denoted as a deterrence effect.

The underlying assumption of obtaining the expected deterrence effect by preventative audits are as follows:
- It provides assistance and education to taxpayers about how to fill in the tax return correctly
- It can increase the perception of the possibility of being audited.

Effect chain:



STEP 3 – Choosing a research design
STA's "Preventative Audits" program is based on random audits. The target population contains around 200 SMEs (sole proprietors or partners in trading companies) who filed late in 2009 within a certain region of Sweden. The treatment group is randomly selected and contains around 100 SMEs. The rest is used as a control group.

The research design for measuring the outcome in term of correct, complete and on-time filing is a 'Randomized Control Trial'. The effect is measured by comparing the average of outcome (i.e., declaration date, declared income, risk scores) of the treated group with the control group. The observed differences in average outcomes between the treated and the control group are interpreted as the effect of implementing "preventative audits".

A restriction of the research design is the small sample size that leads to large statistical uncertainty.

STEP 4 – Ensuring data availability
All necessary data needed for this study were available in STA's internal database. To begin with, the analyst identified the target population by using the "Base Register" for firm type and geographic region and the "Taxation register" for the declaration date in 2009. Thereafter, a randomized selection of the treatment group was performed.

For each of the SMEs in the target population, the following data for tax year 2009 and 2010 was collected from STA's Taxation register:
- Declaration date
- Declared taxable income
- Declared turnover and cost

An audit memorandum was prepared for the tax auditors to fill in relevant information collected during the audits, regarding e.g. the mistakes discovered.

STEP 5 – Securing necessary resources

Staff resources
For implementing the preventative audits, five tax auditors were needed, and working hours were planned to be 500 hours in total.

For the outcome measurement of STA's preventative audits, two type of resources were needed: human resources and equipment. The team involved in the analysis included a coordinator and two analysts with different expertise. The coordinator worked with project management and served as a liaison between the tax auditor and the analysts. The analysts had relevant knowledge in statistics, econometrics and policy evaluation. They were also familiar with the STA's database and were able to work with SQL to prepare the dataset for the analysis.

The equipment required in term of software refers to: SAS, Excel, R (open source free to use).

Timeframe
January 2010 – Identify the target population and select the treatment group
February – April 2010 – Perform the preventative audits
August – December 2010 – Perform outcome measurement
January 2011 – Deliver the report

STEP 6 – Implementing the measured action
During the implementation of the preventative audits, the auditors were instructed carefully by the program coordinator that they had to report the process and findings of the preventative audits and note any dropouts, and that they had to follow the same audit routine, as well as fill in the audit memorandum properly. In this research, 10 dropouts were noted during the implementation. Reasons for dropouts were also noted for later analysis in order to check whether validity of this randomized control trial is threatened as those who completed the process may differ from dropouts in a systematic way.

STEP 7 – Analyzing the data
The analysis started with a description of the audit results by examining the audit memorandum. It appeared that the majority of the SMEs made mistakes in their declarations, for example, putting private expenses in the business accounting records and/or having not filled the declaration form completely. However, the tax administration didn't make additional assessments since the audits were made before filing. The SMEs instead were given an opportunity to correct the mistake(s) themselves.

Next, the analysts measured the outcome of the preventative audits in term of improvement in filing on time, correct filing and increase in tax paid. Thanks to the design of the randomized control trial, the effect of the 'preventative audits' could be

measured by comparing the outcome of the treated group with the control group. Statistical tests were also performed to verify whether the effects found were statistically significant. It appeared that preventative auditing has a positive effect on correct filing. More than half of the audited firms that made mistakes in their declaration forms made the necessary corrections. More importantly, the treated group on average was less likely to be selected for "Risk-based audits" compared to the control group after treatment.

However, no statistically significant effect was found for filing on-time and increase in tax paid. This may be due to the fact that the sample size was too small to make reliable statistical inferences.

STEP 8 – Conclusions and recommendations
Since this outcome measurement showed that preventative audits have a positive effect on inducing correct filing, these findings were used as an empirical support for STA to develop two new instruments:

- Control of documentation obligation. Subsequent legislative changes gives the Tax Agency the right to perform a similar examination as 'preventative audit' without the need of an 'Auditing Warrant'. This type of control does not require to be carried out by an auditor.
- Control of newly started businesses as well as limited companies who opt out of hiring accounting firms to prepare their tax declarations.

By using these two new preventative instruments, STA expects to induce correct filing and thereby improve voluntary compliance with relatively lower costs than performing repressive audits after filing.

STEP 9 – Communicating the lessons learned
The findings of the experiment were first communicated to the head of the auditing function.

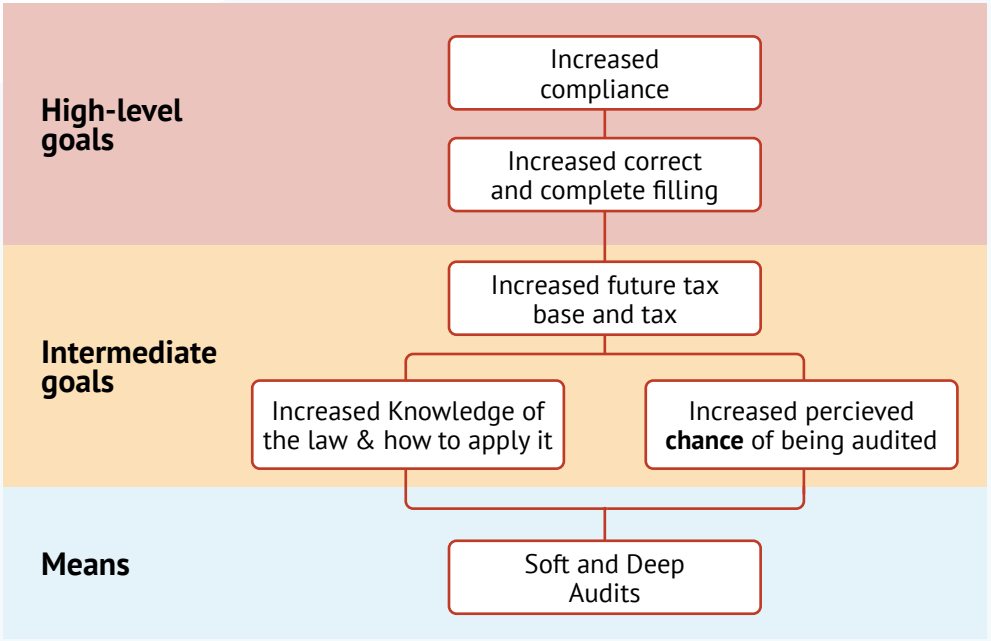
Then the results and the practical conclusions and suggestions for next steps and how to utilize the results through new instruments were presented to the management of the tax administration.

The evidence gathered in the experiment was also presented to the policymakers to support a legislative change towards less rigid modes of audit.

6.2. IRA Example

STEP 1 – Objective setting[36]
A tax audit is one of the instruments in the Compliance Risk Management (CRM) strategy applied by Italian Revenue Agency (IRA) to tackle tax evasion and to promote compliance. Each year, IRA launches a risk based audits campaign on self-employed and small businesses (sole proprietors).

With this campaign, in which different types of audits are deployed, IRA aims:
3. to redress non-compliance and collect the revenue involved
4. to increase compliance (correct and complete filing) of audited taxpayers[37]

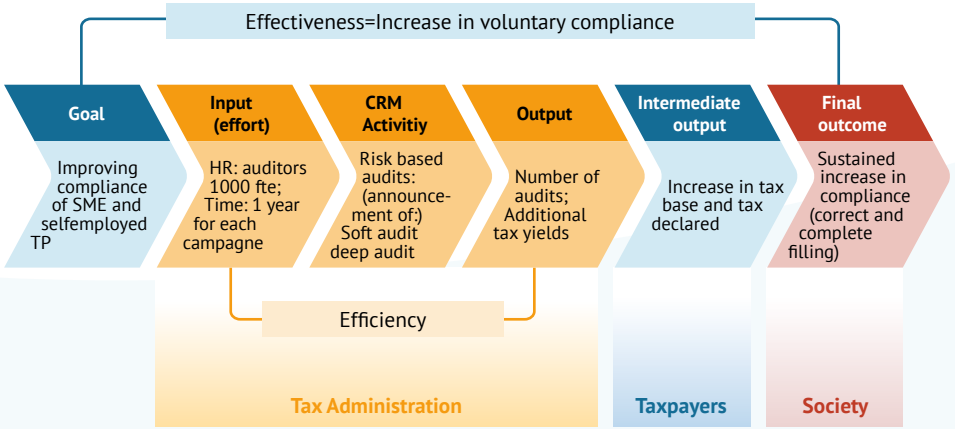


[36] The running example is based upon a practical example of the Italian Revenue Agency (IRA), which is modified for educational purposes.
[37] A third goal, an increase in correct and complete filing of non-audited taxpayers, is not treated in this example. The findings can be found in the paper "Tax compliance and the role of the consultant, evidence from an Italian experience" Alfonso Carfora, Elena D'Agosto and Stefano Pisani

STEP 2 – Forming a hypothesis
IRA expects a positive compliance effect because it assumes that during the auditing process the taxpayer learns about the right interpretation and application of fiscal law. Furthermore IRA expects a compliance effect because it assumes that audits increase 'audit-awareness' and with that, the perception of getting "caught".

IRA takes an increase in tax base and tax declared as a proxy of an improvement of compliance (correct and complete filing).

Effect chain :



STEP 3 – Choosing a research design
The design of the research had to deal with the risk based character of the audits as well as the several years processing time the audits took.

The targeted (treatment) groups were those taxpayers who had undergone different types of audits (soft audits or deep audits or sometimes both). The control group had not undergone any form of audit, but was selected in a way that made them similar to the targeted groups.

IRA decided to choose a difference in differences approach (DID) in order to deal with the selection bias issues induced by non-exogenously assigned audits. According to the IRA auditing policies, taxpayers are selected for audits based on their compliance outcomes, which means that there are pre-existing differences between the audited and non-audited that can influence the outcome after audits. The DID approach allows accounting for those observable and unobservable differences between the two groups of taxpayers under the basic assumption that these differences do not vary over time.

Information on available treatment and control groups was before and after the audits occurred, in a panel data structure.

Restrictions of the research design:
- Some taxpayers could be audited in more than one year, which could drive them to behave differently from those who were audited only once. This could increase or decrease the impact determined by the first audit.
- The analysis did not include information about taxpayers audited before the risk-based audit campaign started. This could slightly affect the control group and the results of the estimates.
- The analysis does not include the quality of the audit as perceived by the tax-payer.
- The analysis did not distinguish for any status of audit as ongoing or concluded.

STEP 4 – Ensuring data availability
The dataset was drawn from two internal sources:
1. Tax audit records
2. Tax returns register

From the Tax returns register, all information on Regional Tax on business activities (IRAP), VAT and PIT was included in the research. From the Tax audit records, all information from the audit records was also included in the research.

The research was conducted on a 10% sample of the reference population of self-employed and small businesses (about 750.000 taxpayers), audited and non-audited taxpayers included.

The computer scientist verified the quality of the data and its format being suitable for subsequent exploration. During this verification process simple mistakes, outliers and incoherencies were discovered.

STEP 5 – Securing necessary resources

Staff resources

IRA outcome evaluation encompassed two types of resources: human resources and equipment. The full team involved in the analysis phase included: a computer scientist and three economic analysts with different expertise. The computer scientist was necessary for data preparation to make the dataset for the subsequent analysis available. The economic experts were acquired for their skills in Applied Econometrics and policy evaluation, one tax auditor and one tax expert for their business knowledge.

The software used for running the outcome evaluation consisted of software for econometric and other statistical analysis: e.g. SPSS, Stata, R (available free of charge on the internet).

Timeframe

It took approximately two years from the start of the evaluation to the final report.

STEP 6 – Implementing the measured action

IRA's risk-based audit campaign has been conducted on a risk-based selection of self-employed and small businesses. Either the local or the central Office was in charge of making decisions about numbers and types of audits. Auditors made their investigations over their jurisdictional area. All auditing results were reported in a central tax audit register. After the campaign a further quality check of available data was conducted to verify outliers and consistency of all information acquired.

STEP 7 – Analyzing the data

Descriptive statistics were drawn to get a feeling of the business characteristics in order to better understand the treatment group and the control group. This was helpful to better interpret the outcome measurement and the possible differences between the treatment and control group.

Next, the analysts carried out the actual evaluation by applying econometric techniques suitable for a proper evaluation estimation. According to the data availability, a difference in differences approach was applied. In particular five outcome variables were examined to understand taxpayers' behaviour with respect to various taxes: IRAP base and tax declared, VAT turnover, VAT declared, PIT declared. IRA compared the trend of these variables in the treatment group with the control group.

The assumption was that the trend in both groups, before audit, would be the same so differences occurring after audits could only be caused by the audits. This crucial assumption for the validity of the estimates is called common trend assumption i.e. the condition for the validity of the estimates requires that the pre-audit trend of the audited taxpayers is the same as the trend of the non-audited taxpayers group.

STEP 8 – Conclusions and recommendations

IRA findings suggest an increase in the compliance proxies (tax base and tax declared) used. The magnitude of the effect differs across taxes; an audit increases the regional business tax declared by 1.8%, the VAT by 5.3%, and the PIT by 5.2%. Likewise, the persistence of the effect of the audit over time also differs across types of tax. While for both regional business tax and PIT the effect remains positive also in the first year after the audit, in the case of VAT this effect seems to vanish in the first year following the audit.

Deep audits turn out to be the most effective action to increase tax compliance. In particular, taxpayers who experienced a deep audit increased the regional business tax they declared by about 19% in the year of the audit, the VAT declared by about 18% and the PIT by 14.7%. Soft audits seem to exert a weaker effect on tax compliance. However, as they often serve as input to subsequent deep audits, their effect may be underestimated.

Based on these results of the research, the administration was recommended to continue with the annual campaign.
Further on the findings were made available for:

- discussion with offices involved in taxpayers' selection rules;
- the next audit campaigns measurements;
- the audit campaigns measurements at different territorial level with a view to improving the knowledge about audits by Local offices.

It was also recommended to compare these results with those from others instruments, i.e. letters or other preventative instruments.

STEP 9 – Communicating the lessons learned

The results were presented to the top management, including the Directors of the IRA Departments, to the audit operational unit within the IRA and were published in the Public Finance Review.

6.3. List of participants

The following representatives of tax administrations participated in the Fiscalis Project Group on Identifying new and innovative methods to measure outcomes in tax administration, chaired by representatives of the Coordination Sub Group of the Compliance Risk Management Platform, Keith Walsh and Lisette van der Hel :

Member State	Participant
Austria	Herbert Mikulasek
Belgium	Dianne Van Den Cruyce
Bulgaria	Vesela Kostadinova
Finland	Maija Keskinen
Ireland	Katie Ryan
Ireland	Keith Walsh
Italy	Elena D'Agosto
Italy	Andrea Spingola
Lithuania	Vytenis Zaskevicius
Spain	Jimena Acedo
Spain	Olalla Alonso
Sweden	Yuwei Zhao de Gosson de Varennes
The Netherlands	Herman Hoorweg
The Netherlands	Lisette van der Hel – van Dijk

This is a publication from:
The EU Compliance Risk Management Platform
June 2020