GOBIERNO DE ESPAÑA
MINISTERIO DE POLÍTICA TERRITORIAL Y FUNCIÓN PÚBLICA
SECRETARÍA DE ESTADO DE POLÍTICA TERRITORIAL Y FUNCIÓN PÚBLICA
INSTITUTO PARA LA EVALUACIÓN DE POLÍTICAS PÚBLICAS

GUIDELINES FOR THE OUTCOME AND IMPACT EVALUATION OF PUBLIC POLICIES

INSTITUTE FOR THE EVALUATION OF PUBLIC POLICIES

2020

GOBIERNO DE ESPAÑA
MINISTERIO DE POLÍTICA TERRITORIAL Y FUNCIÓN PÚBLICA
SECRETARÍA DE ESTADO DE POLÍTICA TERRITORIAL Y FUNCIÓN PÚBLICA
INSTITUTO PARA LA EVALUACIÓN DE POLÍTICAS PÚBLICAS

AUTHOR:

- Javier Rico

Director: Celia Casillas

Guidelines for the Outcome and Impact Evaluation of Public Policies

Secretariat of State for Territorial Policy and the Civil Service

Publisher: Ministry of Territorial Policy and the Civil Service 2020

General catalogue of official publications:
http://publicacionesoficiales.boe.es

NIPO: 277200121

Contents

INTRODUCTION

A neighbouring country initiated an extensive programme of subsidies and financial aid in an effort to lift certain vulnerable collectives out of poverty. The programme was deemed a success: the procedures for granting aid had been correctly implemented and without delays, and the target population also received the grants. As a matter of fact, the coverage rate was 100% of the allocated budget, and the programme was correctly implemented by the organisation in charge, the Institute for Fighting Poverty. The policy makers and managers were congratulated for their efforts. The media also discussed the success of the programme and widely broadcast the news.

But within two years, the majority of the population that had been given aid under the programme found themselves in the same situation of relative poverty.

An analysis of this hypothetical programme could not provide an answer to a key question: Was the programme successful in fighting poverty in said collective? In what measure? What were the real effects of the programme on the section of the population that lived below the poverty line, and on society as a whole? Had the programme objectives really been achieved? These are some of the questions that an outcome evaluation seeks to answer. It is a specific type of evaluation that examines the end-outcomes and effects of a public intervention, plan, policy or programme, once implemented.

The aforementioned example is an approximate representation of how evaluation has evolved as a discipline. In its initial stages, evaluation focused on the execution and implementation of public programmes and policies, their inputs and outputs. In this initial phase, (the 50s and 60s of the 20th century), public policy analysis was of a more prospective nature, given the extent to which it focused on the analysis or evaluation of alternatives prior to the establishment of the programmes. While the focus of attention would shift very soon (in the 60s, given the surge in social programmes in the US, such as the fight against poverty) to the outcome and impact of public policies, it was not until the 70s that it became common practice to perform a retrospective analysis of the generated effects. It was intended to answer the question: What happened? Or: Did we get the awaited outcome? This change in focus helped to overcome some of the limitations of the traditional approaches to evaluation, given that the successful implementation of an intervention does not necessarily translate into real and tangible outcomes.

While public policy evaluation has established the outcome of the interventions as the key element, the development of different tools and theories have contributed greatly to highlighting the causal mechanisms that lead to the effects or outcome. Within the field of evaluation, a noteworthy approach is that based on programme theory and its variants (H.T. Chen, M. Lipsey or C.H. Weiss), where the key element is the logical framework or theory that allows us to explain the outcome and the causal chains that link and explain the how and why of the outcome.

The Institute for the Evaluation of Public Policies, as a body of the General State Administration Services to promote the culture of public policy evaluation and the formulation and dissemination of methodologies for evaluation[1], has sought to facilitate the use of tools that help to assess any policy plan or programme and contribute to the institutionalisation of evaluation, its integration in public administration from the planning stage onwards. These tools include the drafting of specific methodological guidelines on the different dimensions of a comprehensive assessment of public policies, aimed both at evaluators and the managers or public officials in charge of commissioning said evaluations.

The Institute for the Evaluation of Public Policies has also published the following guidelines: Guidelines for the Evaluability Assessment of Public Policies, Guidelines for the Design Evaluation of Public Policies and Guidelines for the Implementation Evaluation of Public Policies. All the guides are published on the website of the Ministry of Territorial Policy and the Civil Service in the Institute section: https://www.mptfp.gob.es/portal/funcionpublica/evaluacion-politicas-publicas/Guiasevaluacion.html.

The goal of these Guidelines for the Outcome and Impact Evaluation of Public Policies is to offer some simple guidelines for assessing the outcome of public interventions with two different target groups. On one hand, we have the persons or bodies in charge and the managers of the interventions who may thus have a general overview of the dimensions and contents of the outcome evaluation. And on the other hand, the evaluators of the public programmes or policies who will find in these Guidelines a suitable conceptual framework, techniques and ways of approaching the outcome evaluation.

The first part of this document, “General Questions” seeks to provide a simple response to basic questions on outcome evaluation, in a question and answer format: what it is, what are the basic concepts and approaches to an outcome evaluation, why is it performed, what does it consist of, and how to perform an outcome evaluation.

The second part, “Methodology of Outcome Evaluation” delves into the details of the analyses required to perform an outcome evaluation. Finally, it examines the evaluation criteria and questions as well as the tools and techniques that may be used in these evaluations.

[1] Royal Decree 307/2020 of 11 February which outlines the basic structure of the Ministry of Territorial Policy and the Civil Service. Article 2.5.

PART ONE. GENERAL QUESTIONS

What is outcome evaluation?

Outcome evaluation is a type of evaluation aimed or centred in a general sense on the effects, outcome, and effectiveness of public interventions. Outcome evaluation has occupied a privileged place in the field of evaluation. This is due to several factors.

On one hand, the increasing demand for tangible outcomes in government interventions, which requires improved performance to combat the deficiencies in providing public services and resources, and to solve public problems. The demand for greater accountability in advanced societies where transparency in public action is a requirement for its legitimacy, has also led to greater focus on this topic. Similarly, there has been a paradigm shift in public management, which has progressed from a management system based on rules and procedures to one that pays greater attention to the outcomes. Terms such as outcome management, results-based management, or performance management have become common expressions, even included in legislation.

The need to adjust budgetary frameworks and public spending constitutes another basic element of the need to provide proof of the positive effects of the action in question, and outcome evaluation is critical to this aspect. Finally, there are two aspects related to evaluation in social science research: the inherent limitations of evaluations that are focused exclusively on the process or implementation; and its nature as a discipline characterised by its practical applicability and the practicality of its research framework, i.e., public policies. This has led to the evolution of research focuses, paradigms and models in a relatively short period of time. In this regard, evaluation is a discipline closely linked to public policy analysis, from which the former inherits its epistemological debates, the tensions between diverging paradigms, and the constant discussion on the main approaches and theories of evaluation. This is a controversial and unresolved question. Scientific theories are not neutral, given that they "frame or endorse a certain worldview" (A. Roth, 2008). This clearly influences how we approach an outcome evaluation, its process, and the methodological approach to be adopted.

This predicament faced by evaluation in recent years is partially due to this demand for publishing the generated outcomes, a characteristic of outcome evaluation. As a matter of fact, they have increasingly made their way into government and legal functions where, for example, audit authorities do not limit their functions to solely monitor the legality and correct use of public funds.

The general definitions of evaluation also contain references to the concepts of outcome, effects, effectiveness and impact. This is undoubtedly why there are definitions with different approaches, models, or types of evaluation, and that contain the core or basic questions of outcome evaluation. In spite of this, the terminology and concepts regarding outcomes

and impacts may be deemed, in line with B. Belcher and M. Palenberg (2018) as deeply unsatisfactory, ambiguous, lacking conceptual clarity and precision. The OECD provides a broad definition of evaluation as "the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability" (OECD, 2010).

A characteristic feature of the OECD's formulation, and other general definitions of evaluation is the inclusion of the term "objectives", which is missing from other definitions such as the one proposed by AEVAL (2015), for whom evaluation is the "systematic and reasoned process of generating knowledge (...) aimed at a comprehensive understanding of a public intervention – be it a policy, plan, programme or standard-, in order to make an evidence-based assessment of its design, implementation and outcome (results and impact)". A definition that is somewhat similar, with regard to the process of creation of public policies and the absence of the term objectives, to the one drafted by Rossi, Lipsey and Freeman (2003) for whom, "evaluation is the systematic application of social research methods to assess the conceptualisation and design, the execution and utility of social intervention programmes". This conceptual difference has important implications with regard to the methodological approach to evaluation in general, and outcome evaluation in particular.

Keeping in mind the OECD's definition, outcome evaluation may be characterised as focusing specifically on the achievement of objectives, their effects, utility, analysis of the effectiveness and the impact -terms that are not synonymous, as we shall see later- and it is performed once the intervention is implemented (ex-post).

The different methods of outcome evaluation have evolved over time and may be grouped into two approaches that summarise the core elements of outcome evaluation: a logical and sequential perspective of the programme which follows the causal chain of results; and another broad focus that centres on or begins with the change produced. Both approaches are present in the different basic concepts on outcomes, as well as in the process to be followed for outcome evaluation, as we shall see in the second part[2].

[2] Another approach eliminates the possible sequences mentioned and the evaluation follows the causal relationship, moving to the top and the bottom of the causal chain.

Fundamental concepts in outcome evaluation

A. The "objectives" of the intervention

An essential concept in outcome evaluation is that of the "objectives of the intervention", to the point that both general and specific definitions of outcome evaluation establish a direct link between the outcomes and the objectives: outcome evaluation therefore, focuses on the degree to which the expected objectives of the programme or policy have been achieved, also linking them to effectiveness. Therefore, alongside the definition established by the OECD, the Centers for Disease Control and Prevention (2007) consider outcome/effectiveness evaluation to be that which "measures program effects in the target population by assessing the progress in the outcomes or outcome objectives that the program is to achieve". For the Office of the Auditor General of Canada (2009), effectiveness evaluation is a tool that uses systematic research methods from different disciplines to evaluate the degree to which a programme is achieving its objectives.

This identification is so wide-ranging that it has important implications for the focus and methodology of evaluation. Evaluation should be centred and focused on the objectives of the intervention and its fulfilment. It must analyse what were the initial objectives, if they were explicit, and compare them to the achievements once the intervention has been implemented. Thus, Tyler, when analysing the degree of fulfilment of the objectives, considers that evaluation must establish the general targets to be achieved, the objectives that lead to them, and compare the outcome with said targets.

This evaluation approach has the advantage of measuring and quantifying the achieved objectives and their attainment with regard to that which was defined as necessary to solve the problem. To which must be added the "how" and the "why" of their achievement, if we are to have satisfactory evaluations. Evaluations based on objectives define their framework of evaluation and focus their evaluation efforts and resources. But this approach also has evident limitations. It excludes from the evaluation all outcomes that are not related to the objectives of the intervention. It is equally important to examine undesired outcomes as it is to analyse the outcomes related to the objectives. The evaluations that focus on the objectives or that adopt programme models can (and tend to) inevitably underestimate the importance of these aspects, thus running the risk of being unable to include potential effects (and the causal chains that provoke them) in the models, as well as other routes that are not expected by the programme theory (E. J. Davidson, 2004).

Additionally, the term "objectives" is under no circumstances neutral, and there may be clear differences between what is considered an "official" objective and a real objective. As a matter of fact, public problems are neither neutral nor are they completely objective in nature, rather they are social and political constructs. As are the formalised objectives of public interventions.

Another problem of evaluation centred on objectives is related to the inexistence of formal objectives where the declared objectives are not the real ones, and to approaching large-scale public policies. Finally, and depending on how the objective-based outcome evaluation is performed, questions linked to contribution and attribution, and the causality of what is achieved and whether it was a consequence of the public policy or programme. We shall study these issues further over the course of these Guidelines.

These aspects are overcome or corrected using a free-goal evaluation approach. Initially formulated by Scriven, this type of evaluation does not focus on the objectives formulated during the intervention rather it takes a comprehensive approach to the object of evaluation, where the central issue is the changes to reality, beyond the hypothetical planning of the intervention based on certain declared objectives. It thus overcomes what may be called tunnel vision, given that the evaluation does not focus exclusively on certain elements or standards against which to evaluate (the objectives) but covers all the changes and outcomes that take place, not just those related to the objectives.

Additionally, contrary to a vision focused on a target population of the objectives and certain sought or desired "objectives", a free-goal evaluation makes it possible to analyse the changes that take place in the population that is not the intended beneficiary but experiences direct as well as indirect changes and effects. It also lets us analyse changes in outcomes that have not been sought by the intervention. As a counterpoint, they are usually more complex, and carry the risk of evaluating questions that are not central or do not cover all aspects, as well as leading to difficulties in focusing the critical aspects of the evaluation.

B. The "outcome" of the intervention

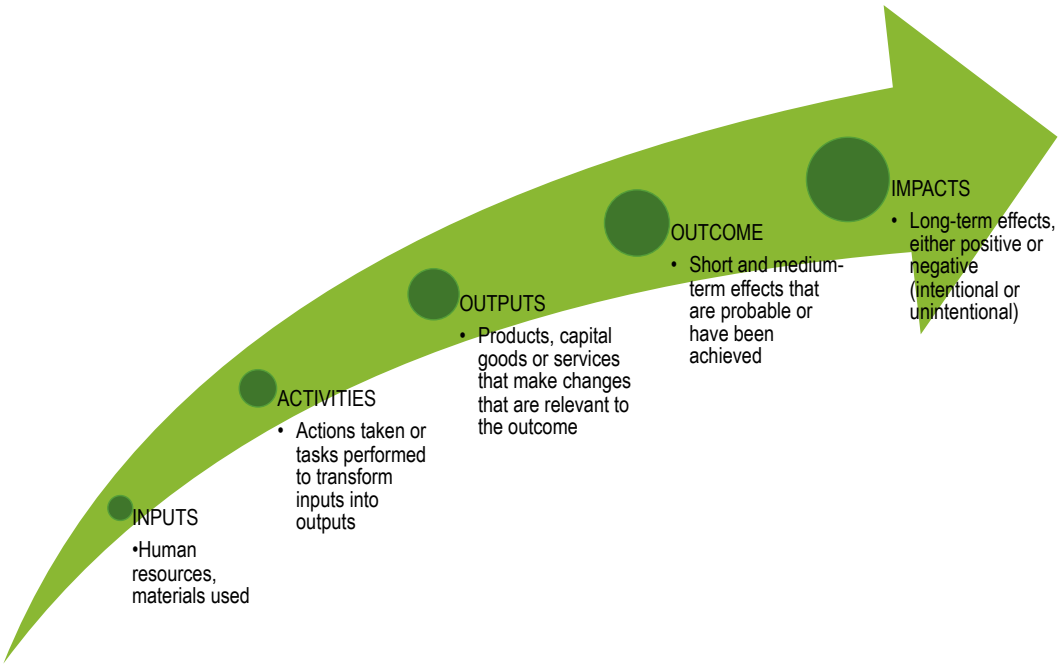
This is the fundamental concept that defines outcome evaluation. One of the most comprehensive definitions is that of the OECD for whom outcome is "the likely or achieved short-term and medium-term effects of an intervention's outputs", as opposed to impacts, which are medium and long-term effects.

This definition, along with similar ones formulated by evaluation associations and multiple authors, lay down certain defining elements of the concept of outcomes:

- Their time-based nature, as they may be either short-term or long-term outcomes
- They arise from the "outputs" of the intervention and are subsequent to them
- They may be initial projections or actively come to pass
- The causal link is centred on the intervention, i.e., the outcomes are derived directly from the intervention outputs
- Outcomes may be intentional or unintentional and positive and/or negative. However these aspects are not dealt with clearly in the definition, in comparison to the OECD's definition of impact.

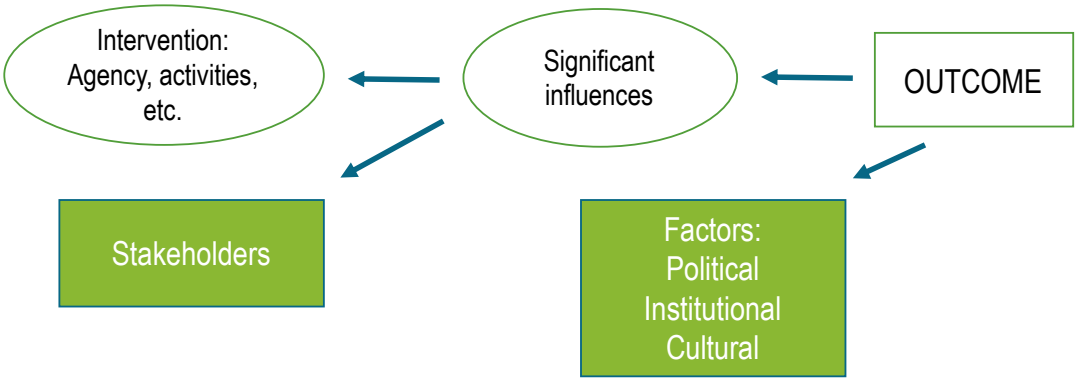
The focus is on the intervention and the logical sequence of inputs, activities, outputs, outcomes and impacts. Thus, the outcomes are the consequence of the programme or policy outputs, which in turn are derived from the activities or tasks performed to transform different inputs into outputs. This perspective lends greater weight to the analysis of the outcomes in relation to the objectives initially set by the programme.

Figure 1. Results chain. Source: Author's own.



The focus of evaluation has evolved towards another approach that centres on the change produced, on the modification or alteration of the existing conditions or situation, and how the outcome of the intervention contributes to this change. The perspective of the analysis is not so much on the formal or implicitly declared objectives as on whether progress or an outcome has been achieved and how, why, and what circumstances have changed the outcome or the status, given that many other factors play a role alongside the intervention, programme or policy. Thus, "the intended result is the specific dimension of well-being and progress for people that motivates policy action, i.e. what is intended to be changed, with the contribution of the interventions designed" (EU, 2014). Or the degree or scale of something that was expected to change as a result of the intervention, either in the participants or in their environment. Change is much more relevant in this approach where outcome analysis commences not with the intervention but with the change effected, following a retrospective logic.

Figure 2. Changes effected. Source: Author's own based on UNDP Evaluation Office (2002)



We are therefore dealing with two main perspectives that are independent of how the cause is measured or defined (B. Belcher and M. Palenberg, 2018). A perspective from the intervention, which describes the changes that it leads to. The intervention is the origin of the causal system and focuses on the part of the change it produces. The other perspective is *systemic*, where the point of view shifts from the intervention to the changes of interest. Beginning at the changes, it looks back at the contributing causes. Some may be related to the intervention, while others may not. In this regard, the existence and degree of influence of the intervention in the changes continues to be of central interest, but this does not necessarily mean a total or even partial attribution.

In both cases, one question linked to the results chain is the position of the outputs, which are the primary objects obtained by the intervention. Are outputs an "outcome"? Or do they occupy an earlier stage? Some proposals (such as that of the OECD) include them within the outcome. Nevertheless, other authors consider them to be part of the execution or implementation of the intervention. In fact, an output is not an outcome per se. Let us take a healthcare policy or programme that seeks to reduce waiting lists or improve healthcare for a population. The increased number of hospital beds is not an outcome. It is an output of the programme that will potentially contribute to reducing waiting lists alongside other outputs, but in and of itself, it does not improve waiting times or healthcare. Or when we look at an intervention that seeks to enhance training and the acquisition of skills and abilities in government employees by means of training activities. The courses attended by the employees are not an outcome but an output that, under certain circumstances, may become an outcome, but does not necessarily translate into improved performance or skills (see AEVAL evaluation, 2013).

The term outcome[3] has an evident polysemiotic nature:

- Outcome, with regard to performance. That which is produced or obtained with the resources allocated to an intervention

[3] One of the difficulties inherent in the Spanish version of this term is that they are synonyms, whereas in English they are different. The terms "result" and "outcome" are clearly differentiated in English, but are synonymous in Spanish.

- Outcome, with regard to effects. The immediate changes derived from an intervention.
- They are the benefits or changes produced in the target population during or after the intervention
- Outcome, with regard to impact. Long-term, direct and indirect, positive and negative socio-economic effects that may be observed after a certain period of time after performing an intervention
- Outcome corresponds to the "gross effects" of an intervention, in contrast to the impact, from one possible perspective where the latter measures the net effects

Outcome is closely linked to other basic concepts of outcome evaluation, such as objectives, effectiveness, and impact.

A relevant aspect of outcome evaluation is not only the knowledge of its scope or its achievement, but also determining how and why these changes have been produced or not. This may constitute a limitation of some approaches to this type of evaluation. Another inherent limitation is multi-causal effects: it is possible that the intervention per se has not contributed to the total achievement of the outcome, which may respond to other factors.

C. The "effectiveness" of the intervention

As mentioned before, the different definitions of evaluation generally include the term effectiveness. But what is meant by an effective public action[4]? The concept of effectiveness may be considered, as indeed is stated in numerous definitions, for example in the Glossary of Key Terms in Evaluation and Outcomes-Based Management of the OECD, as a synonym for achieving targets and objectives. Thus, an effective intervention would be one that achieves its (strategic or operational) targets and objectives that it aimed for. Both terms may be related in evaluation. Thus the term effectiveness may be deemed as "the extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance" (OECD, 2010).

Effectiveness may also be deemed the degree to which a programme or policy yields the desired effect, focusing more on what the programme really pursued in connection with the public problem that motivated it. This dual meaning is referred to, among others, by AEVAL when defining the evaluation criteria, in defining effectiveness as "the extent to which the problem, demand or requirement behind the motivation has been solved. Effectiveness also refers to the degree to which the objectives stated in the intervention have been achieved. They must be mentioned with regard to the awaited outcome of a programme, when the latter is explicit. Or based on the evaluators' identification of the real underlying objectives, when implicit. "(...) On the other hand, effectiveness may have multiple dimensions: economic dimension, social dimension, etc." (AEVAL, 2015). Nevertheless, effectiveness is intrinsically linked to solving a public problem and not solely to fulfilling certain objectives.

[4] In English, the term effectiveness may have multiple meanings: efficacy, effectivity, and effect.

Effectiveness (or effectivity) is the degree to which an intervention has achieved the expected outcome, that is to say, its specific objectives (intermediate outcome). The theory of change may be implicitly considered, as effectiveness allows us to establish the link between the primary outputs of the intervention (products, services or outputs), its intermediate outcome, and its final outcome. Therefore, a policy or programme is considered effective when the outputs achieve the awaited outcome.

Effectiveness is sometimes compared with impacts, where it focuses exclusively on the sought-after outcome, excluding long-term and unwanted outcomes. The problem of this approach from the point of view of evaluation criteria is that there would be no specific criteria that focuses on the impacts.

D. The "impact" of the intervention

Once again, different meanings are attributed to the concept of "impact", leading to a degree of terminological confusion. Thus, impact may be:

- The ***long-term*** outcome (either primary or secondary, directly or indirectly produced by an intervention, either intentionally or unintentionally). Impact here contrasts with "outcome" which would be limited to short and medium-term achievements (OECD definition).

- For other authors, the term impact is related to the changes or effects produced in ***society in general***, beyond the boundaries of the direct targets of the intervention. It deals with all sorts of effects, sought and unsought, positive and negative. It contrasts with outcome as the former refers to the final consequences for the target population or the direct beneficiaries of the programme.

- Finally, the term impact refers to the ***net effects*** of the intervention. In contrast, outcome refer to the "gross effects". Given that the changes in the environment, in the beneficiaries of the intervention, and in the entire society may be due to multiple causes, impact evaluation seeks to reveal the changes that may uniquely and exclusively be attributed to the public programme or policy, once the rest of the factors or incidental causes have been controlled. The literature consulted appears to consider the term "effect" to be equivalent to outcomes due solely to the programme.

The key aspect of "impact" is that it delves into the attribution of the programme and focuses on the causal dimensions (cause and effect). The basic question that is asked is: What is the impact or effect of a programme on its expected outcome? While the question is relevant for different approaches to outcome evaluation, the characteristic of a specific type of impact evaluation (attribution) is to use a counterfactual: what would have happened to the target

population of the programme if the programme had not been implemented or if the targets had not participated in it. It uses control groups and both experimental (randomised controlled trials) and non-experimental methods of analysis. We shall discuss these elements in greater detail later.

Impact evaluations are highly useful when subjecting the obtained outcome to rigorous statistical analysis. They also focus on a basic question of outcome evaluation, attribution, and tend to solve issues of internal validity. They must be performed under optimal conditions for their results to be fully valid. On the other hand, they face a series of difficulties or "disadvantages".

E. Contribution and attribution: the search for a causal relationship

The concepts of contribution and attribution are extremely significant for evaluations in general and especially outcome evaluation. Both concepts are linked to the idea of causal relationships or inference, dealing with the "effects" of public interventions and the causes that provoke the changes.

Contribution is the degree to which an intervention has made certain outcomes possible, has influenced or helped to cause certain observed outcomes, in order to achieve the pursued objectives or the concrete change observed. The contribution may be formulated as evaluation questions that are part of the effectiveness criteria of public interventions: "Has the programme contributed to obtaining the outcome of interest?" *"What is the extent to which the intervention has contributed to achieving the outcome sought?" "To what degree does it contribute to the observed outcome?" Or: "Can the intervention, by itself, have a significant weight or influence on the outcome?"*

One of the characteristic traits of contribution which differentiates it from attribution, is the underlying idea that the programme or policy does not have sole involvement in the outcome achieved. As an analytical focus or strategy, contribution confirms or invalidates the causal assumptions from a chain of reasoning which assumes that the programme has made the observed change and analyses how, why, and to what degree (intensity) has it contributed, always keeping in mind that other factors also contribute to the outcome.

The term **attribution** refers to the fact that the intervention has caused the observed change: determining the extent to which the programme or policy has provoked this change. Or from another position or perspective, the possibility of identifying what part of the change or outcome may be solely and exclusively attributed to the public intervention. "Attribution implies causality and establishing causal links and explanatory conclusions on the relationship between the observed changes, anticipated or unanticipated, and the specific interventions. The study of attribution seeks to determine the extent to which the programme caused the observed results" (Almquist, 2011).

One of the complexities of an attribution analysis is that there are very few occasions when we may speak of exclusive attribution, that is to say, when the outcome or impacts are caused exclusively by the intervention, without purely contextual, economic, social and institutional factors, other public interventions, or the stakeholders themselves having contributed to eliciting them. Attribution is usually joint, when the aforementioned factors combine to produce the change, and their absence may lead to the inability to achieve the outcome or impacts, or not to their full extent. The presence of multiple interventions makes internal validity hard to establish and attribution, hard to demonstrate. Finally, it is worth mentioning other elements that are not accounted for sufficiently in attribution studies, such as alternative or multiple causal pathways, in assumptions where outcomes may be achieved by alternative means, in which case, cost-benefit analysis or the study of the restrictions in achieving the outcome becomes highly relevant. As for the rest, in such a complex world and systems, simple linear causal models do not work.

The difference between the concepts of contribution and attribution may be clearly observed if we consider the dilemma that a change motivated by an intervention is not "attributable" exclusively to it, but "contributes" to it. Also, attribution requires quantitative methods and outcome variables that are perfectly measurable, whereas contribution does not necessarily have these restrictions.

Types of outcome evaluations

There are numerous taxonomies or classifications of outcome evaluations and some revolve around the concepts mentioned before. Schalock (1995) establishes three types of outcome evaluations:

- The evaluation of effectivity or effectiveness. The degree to which the programme obtains its targets and objectives.
- Impact evaluation or analysis. It determines whether the programme makes a difference when compared to its absence or to an alternative programme.
- Cost-benefit analysis. It determines if the programme benefits outweigh its costs.

Another classification is that which hinges on the objectives of the intervention. Thus, we may distinguish between:

- Evaluation focused on the objectives of the intervention.
- Free-goal evaluation

Based on the timetable of the outcome analysis:

- Short-term outcome
- Medium-term outcome
- Long-term outcome or impact

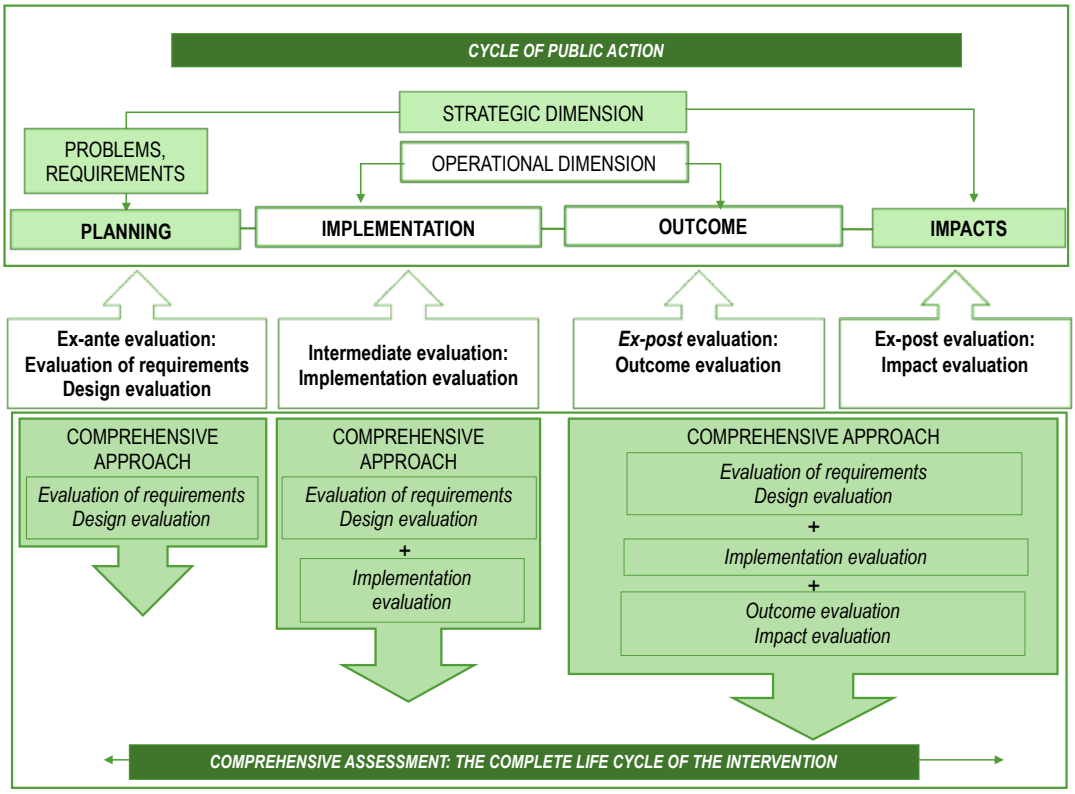
Based on the evaluation target:

- Outcomes within the target or beneficiary population of the intervention
- Within society as a whole

Outcome evaluation within the general framework of evaluations. Its relationship with comprehensive assessment

Although outcome evaluation is a specific type of evaluation, generally focused on the outcome of public interventions, the relevance of considering said evaluation within the wider framework of evaluations or the concept of comprehensive assessment must not be ignored. This approach considers that public policies are action processes characterised by their complexity and inter-connections with public problems, which require an approach that fully includes the problems as well as their solutions as opposed to fragmented and decontextualised analyses. It thus integrates the political-strategic analysis of policies -which includes problems, assessments, stakeholders, intervention theories and hypotheses, formulation and formalisation of the intervention and its impacts- with the analysis of the operational aspects of its rollout, that is to say, its objectives, resources, processes and intermediate and final outcomes (AEVAL, 2015).

Figure 3. Cycle of public action and comprehensive assessment. Source: Author's own based on AEVAL (2015).

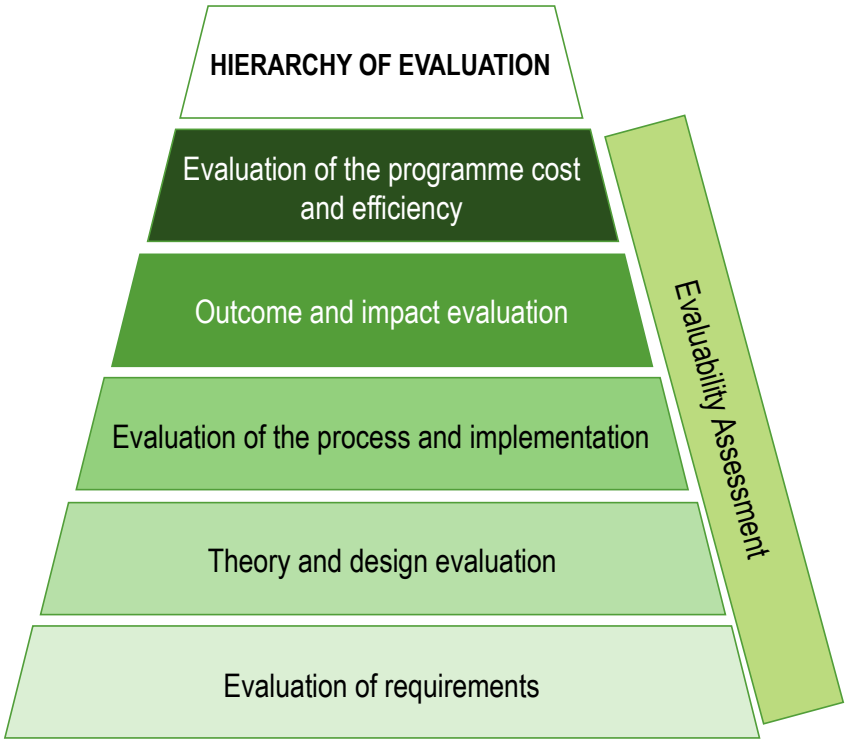


This creates the need for the overall approach of comprehensive assessment where outcome evaluation acquires a full-fledged meaning. The evaluation should not focus on "what" has been achieved, but also on "how" and "why", for it to be of real use.

It is not just about analysing the outcome per se of public programmes and policies, but also the factors or elements that lead to or condition the outcome, or why the projected objectives have been achieved or not. Often outcomes are intrinsically linked to how the public problem that engenders the public action has been diagnosed and defined; the design of the action and the existence of a valid theory of change that correctly identifies the cause-effect relationships. All these aspects constitute the specific field of design evaluation. Design defects are usually one of the key factors to be considered when comprehending and analysing outcomes. The context or framework in which interventions take place, the motivations and interests of the participating stakeholders, and finally, the implementation itself, also aid in understanding the outcomes obtained.

In this regard, the hierarchy of evaluations created by Rossi, Lipsey and Freeman (2003) acquires special significance. They outline five levels, where outcome or impact evaluation occupies the fourth level. There is little point in performing an outcome evaluation when the existing requirements, the intervention design and the implementation process followed have not been clearly identified: "If we know that the social need is properly understood, the program theory for addressing it is reasonable, and the corresponding program activities and services are well implemented, then it may be meaningful to assess program outcomes".

Figure 4. Outcome evaluation within the hierarchy of evaluation. Source: Author's own based on Rossi, Lipsey and Freeman (2003).



What is the role of outcome evaluation?

Outcome evaluation essentially serves to answer the question of ***"what"*** a public intervention has achieved. It asks what effects have been produced as a result of applying the programme or policy and whether the awaited outcome and effects have been achieved. Finally, it accounts for all the improvements experienced by society as a result of the public action.

The utility of this type of evaluation depends on the specific focus adopted by outcome evaluation. A complete impact evaluation delves into not only the expected but also the unexpected outcomes in the target population and in the entire society, and undoubtedly constitutes the most desirable form of outcome evaluation. Both in the short term and the long term. If it has a goal-based approach, it will provide proof of their degree of achievement, whereas a free-goal approach will examine problems with a wider scope. Likewise, counterfactual evaluations or evaluations of the net effect of the interventions allow us to determine only the effects that are exclusively attributable to the intervention.

On the other hand, the use of evaluation is linked to the requirements or needs of the person or body commissioning it. This implies that on certain occasions, they wish to know exclusively whether the established objectives are achieved within a specific timetable, and thus it must be correctly assessed if an in-depth analysis of questions on attribution or contribution of the interventions is to be included, especially if there is limited time to perform the evaluation.

This last element of utility is clearly linked to the "objectives" sought by the outcome evaluation, which may be different, and may vary according to the purpose, time, focus and duration of the evaluation. In its development aid programmes, the United Nations generally establishes four objectives of outcome evaluations, although some of them are specifically aimed at development aid:

- Evaluation of the progress made towards the outcome
- Evaluation of the factors that affect the outcome
- Evaluation of the key contributions made by the intervention (outputs) to the outcome
- Evaluation of the partnership strategy or in general terms, how the intervention operates alongside other interventions or with other relevant stakeholders.

The complete usefulness of outcome evaluation therefore lies in including analyses on the design of the interventions, the context in which they are produced, and the process of their implementation, as mentioned earlier.

The utility of evaluation is also related to the requirements that evaluation seeks to fulfil and the time that has transpired since the appearance of effects, which in turn conditions how an outcome evaluation is performed: whether 6 or 12 months have passed, or one or two years or more since the implementation of the programme or policy.

What does outcome evaluation consist of?

As mentioned earlier, outcome evaluation is the analysis of the different dimensions of a public intervention which allows us to draw conclusions, based on evidence, with regard to the effects provoked by a public intervention. This process must have certain specific methodological elements, of which one of the most commonly used is the evaluation matrix consisting of the evaluation criteria and questions.

1.-Analysis of the outcome of the intervention

It refers to the analyses to be performed on the outcome of the intervention and the causes leading to them. It analyses the intervention dimensions that have influenced the achievement of the outcome from the design stage onwards, the implementation elements and external factors that have influenced them. It seeks to analyse the contribution and attribution of the intervention with regard to the objectives that were defined in the planning stage or in the produced change. With these analyses, it seeks to answer the questions of the evaluation matrix.

2.- Criteria, questions and matrix of the outcome evaluation.

The ***evaluation criteria*** provide benchmarks (yardsticks, standards, principles, etc.) that let us obtain useful information in order to assess the evaluation. In outcome evaluation, the most important criterion is that of effectiveness.
The figure below summarises the criteria that are frequently used in an outcome evaluation.

Figure 5. Evaluation criteria directly related to outcome evaluation. Source: Author's own.

Criteria associated with outcome evaluation	
Effectiveness	Degree to which the problem, demand or requirement responsible for the intervention has been solved
	Degree to which the objectives established in the intervention have been achieved
Efficiency	The extent to the which the desired outcome and/or effects have been obtained with the least amount of required resources
Sustainability	The degree to which the benefits of the intervention continue over time
	"Situation in which net advantages are able to resist risks with the passage of time" (EuropeAid, 2006)

The ***questions*** are the basic unit of research. They may be defined as the queries and hypothesis to be confirmed which allow us to execute the evaluation. The list of questions and their associated criteria, as well as the sources of information, measurement indicators, techniques and tools, are included in the evaluation matrix, which constitutes the tool that logically integrates all these elements. The evaluation criteria and questions are included in the matrix and contain the focus and scope of the evaluation.

3.- Analysis techniques in outcome evaluation

This triangular approach is a requirement of comprehensive assessment, as it considers all the theoretical-scientific perspectives that are considered relevant and useful for evaluation. For this, all types of ***techniques and tools*** are used. These Guidelines mention some of the most commonly used ones.

How is an outcome evaluation performed?

Outcome evaluation is a process that examines the effects of public action, the terms under which these outcomes have been obtained, their causes or factors and the degree to which they are attributable to the intervention.

The process unfolds on the initiative of the manager or the person or body in charge of the intervention, whose decision is usually reflected in an initial document that contains the analysis of the commission. The process is carried out by analysing the listed dimensions and ends with the evaluation report which contains the findings, conclusions, and recommendations. The duration of the evaluation will depend on the complexity of the intervention and its nature, characteristics, and conditions, which include the resources allocated for the evaluation.

The process concludes with an evaluation report, which must describe the result of the investigation, the different analyses performed and the findings obtained, usually following a structure based on the evaluation queries used and their associated criteria. The final report must include a conclusions and recommendations section, always based on the obtained evidence. If we think of evaluation as another public intervention, then we may close the cycle with a follow-up of the evaluation[5].

[5] The recommendations of the aforementioned AEVAL Guide 2015 may be followed when drawing up the report.

PART TWO. METHODOLOGY OF OUTCOME EVALUATION.

There is no single method or process of correctly performing outcome evaluation. The adoption of a certain evaluation perspective or approach conditions how it is performed and the aspects to be considered. Thus for example, adopting a neo-positivist vision involves the predominance of the empirical falsifiability of the hypothesis, the development of generalisable causal models, the intensive use of statistical techniques and data analytics, and the tendency to use experimental techniques, etc. Post-positivist proposals place greater emphasis on comparing methodologies, including discourse and social judgement elements, the use of qualitative techniques, etc. Besides, the target of evaluation essentially conditions the manner in which it is performed: a small project or programme that may produce a series of limited outcomes is not the same as a policy or large-scale programmes. The hypothetical outcome that is sought with the intervention is also different, and on occasion, is limited to a population or a reality with clearly defined profiles, and on other occasions, is wide-ranging (generalised interventions). They may be outcomes sought in individuals or changes in society.

To sum up, outcome evaluation is more complex than other types of evaluation, especially project evaluations or implementation evaluations. It is not easy to unravel and measure outcomes in the public sector and establish clear links between the activities of an intervention and the outcome in terms of cost-effectiveness.

Time limits, data availability also condition the evaluation. When feasible and appropriate, depending on the evaluation approach and objectives, extensive micro databases may be used. In the same way, performing an outcome evaluation in a relatively short period of time after the main outputs of the intervention have been delivered, is not the same as performing one after two or three years.

Therefore, these Guidelines are aimed at determining what is to be analysed for a rigorous outcome evaluation, the steps to be followed, and what is the standard process for performing it, as well as highlighting the critical aspects of an outcome evaluation. While it is true that this type of evaluation must focus on and reference the outcome, and therefore not the implementation processes or activities, nor the analysis of the design depth, it must include key elements of both aspects that may explain or condition the awaited outcome.

The second part is structured in the following manner. Firstly, instructions on how to approach the outcome are established, for which we must always take into account the conceptual framework of the outcome evaluation highlighted in the first part. Secondly, it details certain approaches to the analysis of attribution and contribution, which are key aspects of outcome evaluation. Finally, it details the process to be followed when performing this type of evaluation.

The outcome of a public intervention and how to measure it

The critical aspect in outcome evaluation is logically, the analysis and evaluation of the outcome or changes that are produced in the existing situation. It forces us to consider several aspects. The first of them is what is understood by and what is the scope of the "outcome" of the public intervention. Is it the fulfilment of the objectives? Is it the "change" produced in the status of the target of the intervention? Is it the changes sought or any change produced? In reality, these questions reflect the pervasiveness of the different terms discussed earlier. In any case, the evaluator must clearly establish in the evaluation the specific and concrete meaning of the term outcome of the intervention.

As we have repeated throughout these Guidelines, this depends largely on the target of the evaluation and its scope. In programmes characterised, for example, by a clearly defined target population, it is easier to establish the outcome and impacts. In the case of interventions with more general objectives or targets -especially in large-scale policy evaluation- they may be more difficult to define, given that we are faced with phenomena that have multiple dimensions.

The outcome may also be changes in conduct or behaviour, skills, knowledge, attitudes, values, conditions, different attributes or status, thus their measurement and implications may also be different. Occasionally, the motives behind the action are preventive and the objective is to prevent the occurrence of an event or phenomenon, therefore, maintaining this positive status or the absence of the problem is also an outcome, although hard to measure.

When identifying the outcome in an evaluation, the following critical goals may be highlighted:

1.- Identifying Outcomes

The first **consists** of **identifying the outcome** initially awaited or required to solve the problem, as well as the ones observed, outcomes that must be systematically broken down into their constituent dimensions, aspects or characteristics and analysed from multiple perspectives or approaches, especially identifying their critical characteristics. While there are clear, specific and perfectly objective outcomes, there may not be the case for other phenomena, they may have more vague or ambiguous features and their measurement may be more indirect. Let us take, for example, the outcome of achieving that a certain collective is healthy from a healthcare perspective. The meaning of healthy must be precisely established, as well as its dimensions.

To be noted

In the initial stage of identifying outcomes, it is worth detecting them from a ***broad point of view****, without focusing on the objectives, in order to detect all positive and negative outcomes. In contrast to the predominance of quantitative techniques in outcome evaluation, using qualitative techniques and inductive processes makes more sense here.*

It is also worthwhile to use participatory tools that incorporate the opinions and perceptions of the rest of the stakeholders into the results framework. It must be noted that what constitutes an outcome for the manager or designer of an intervention may not necessarily be so for the programme's target.

When identifying the outcomes sought (or reconstructing them if they have not been formally established), the evaluator's priority should not be whether the outcomes are measurable or not, or even how they may be or were achieved (theory of change or of action), as these are aspects to be studied later.

However, depending on the process followed for the outcome evaluation, which we talk about in detail later, it is worth establishing different levels of outcomes, given that one outcome may contribute or lead to the achievement of another. Additionally, there is no specific number of "outcomes" of an intervention.

A ***pragmatic*** approach between the scope and depth of the evaluation and what may be evaluated should prevail, depending on the available resources (time, evaluator capacity, economic resources).

It is best to avoid ***very long-term*** measurements of outcomes, according to the characteristics and nature of the intervention. Seeking to measure outcomes for a period greater than five years, for example, may fail to take into account the greater impact of other factors in the outcomes. The greater the time of the outcomes, the lower the potential repercussions of the programme. The nearer in time, the higher the probability of the programme's effects.

2.- Identifying key mechanisms

The second goal consists of ***identifying the key mechanisms*** of how and why the outcomes are obtained and especially in cases where they are not obtained. If benefits have not been obtained, the reasons why or the location of the problem must be demonstrated.

The outcomes are conditioned by different elements of the programme that lead finally lead to their achievement. The correct identification of the public problem and the targets, the choice of the solution, and a coherent design that correctly defines the measures and actions to be taken with regard to the set objectives, are key elements for the outcomes. That is, the analysis of criteria such as suitability, internal coherence and complementarity are essential in outcome evaluation. Frequently, poor outcomes are the result of poor design or poor identification of the problem to be solved. Therefore, they must be studied in outcome evaluation, without it being restricted to an analysis that focuses solely and

exclusively on results. This also makes it possible for the evaluation to demonstrate how, and the underlying causes behind the failure to obtain the desired or awaited change, which in turn makes the evaluation highly important as it demonstrates how the programme may be improved from the point of view of its relevance to the problem or its design. Evaluation is essentially a feedback tool that makes it possible to redesign interventions so they may be increasingly effective and efficient. In this regard, comprehensive assessment acquires special importance.

Example of change and outcome measurement. Evaluation of administrative burdens in company creation (AEVAL, 2012)

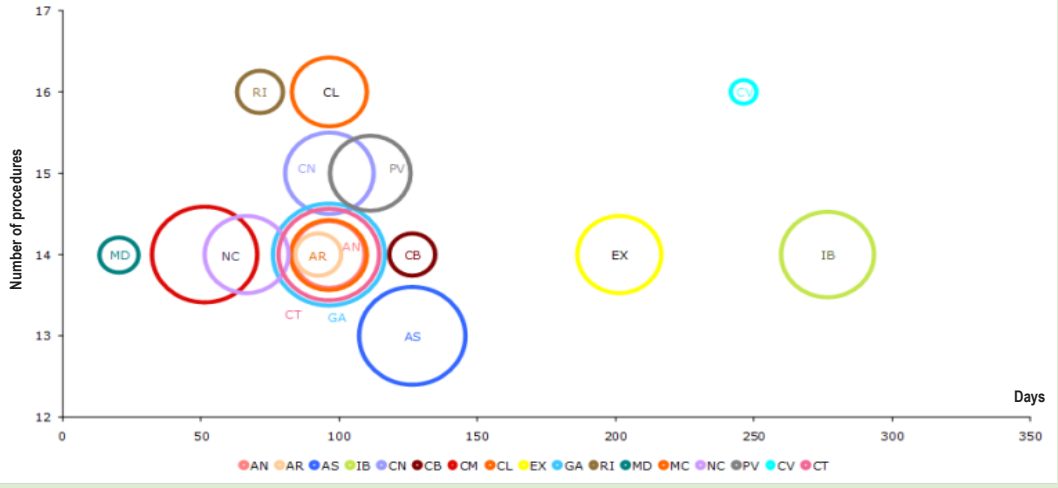
The observed change or outcome was measured on the basis of three outcome indicators: the time taken to establish a company, the number of procedures required to achieve it and the cost of the procedures. It used the following methodology:

-Reviewing the regulations and the requirements for establishing a company, and the current reality of administrative procedures in each Autonomous Region, taking the nation's capital as reference.

-Identifying seven types of companies that cover a wide range of company classes, nature and characteristics, and all the procedures required to set each of them in motion. This included both face-to-face and online procedures.

-Mystery shopping, which consisted of pretending to be an entrepreneur and requesting information on the procedures needed to launch the activity in question in each city.

This helped to achieve real indicators of outcomes, which were compared with those obtained in an ex-ante evaluation. Finally, the contribution and attribution of the measures implemented by the General State Administration and the Autonomous Regions towards reducing all types of costs for launching a business were identified.



The evaluation states that "the situation in 2011 of the process of creation of ICT companies (...) in different Autonomous Regions shows that the majority is situated in the range of 14-16 procedures, with a certain dispersion with regard to the duration of the procedure that may range between 50 and 150 days. If additionally, the cost of the procedure (reflected by the diameter of the circumference) is taken into account, the best global outcome is that obtained by the Region of Madrid, the one closest to the axis and with a small diameter. The regions that are farthest from the axis are Extremadura, Comunitat Valenciana and Illes Balears". On the other hand, "it is evident that the administrative 'costs' are concentrated in municipal activities, in contrast to other state procedures or others whose regulatory development and ordinary management correspond to the Autonomous Regions".

3.- Establishing indicators

A third goal or element is the **establishment of the methods, means and indicators** that let us correctly measure the outcome. The initial conceptual identification of the outcomes needed to solve a problem and the relevant dimensions should not be restricted to those that are initially considered to be measures. The evaluation must explicitly mention the elements that cannot be measured and the reasons why. The evaluator must ask themselves whether the measurement of the outcomes lets them correctly identify the goals of the programme, and the possible shortcomings of the overall measurement. Nevertheless, outcomes may occasionally be intangible and hard to systematically measure. Or they must be measured by means of a more qualitative analysis or by creating proxies or reference groups.

To be noted

*Output or services **indicators should not be confused** with the results. The outputs lead to the outcomes, but they do not constitute the outcomes. On certain occasions, it is possible to clearly distinguish between outcomes and outputs, whereas on others, the difference is not so clear.*

The indicators must reflect the effect, dimension or aspect of the outcomes as exactly and rigorously as possible. Some outcomes are easily observed and measured with relatively simple indicators: for example, a proportion, percentage, absolute values, mean, etc. Other outcomes are not directly measurable or do not capture all aspects of the result.

Indicators must fulfil SMART criteria, they must be: specific, measurable, achievable, relevant and time-bound. In any case, it is worth using indicators that are globally accepted by international organisations, as well as adapting these indicators to the objective and the target group of the programme.

One approach to indicators is that proposed by H. Hatry et. al (1994), where the evaluator may formulate a series of questions on the indicator(s):

- Can we observe and measure this outcome?
- What does this outcome mean exactly?
- Does the indicator tell us whether the outcome has been achieved?
- Is there at least an indicator for each outcome or dimension or aspect of the outcome?
- Was the indicator able to measure any important element of the outcome that had not been measured by another indicator?
- Is the construction of each indicator sufficiently specific?
- Does it indicate the characteristic or change to be produced?

USAID (2010) identifies seven criteria that help in the selection of results indicators: direct (it clearly measures the awaited result); objective (the indicator is not vague about what it measures and the data to be compiled); useful for management; attributable or may be plausibly associated with the programme[6]; practical (the data may be collected within an acceptable period of time and at a reasonable cost); adequate (the indicator or set of indicators is enough to measure the status of the result); and disaggregated, as necessary. The process of selecting and developing indicators may be broken down into a series of stages (USAID):

- Develop a participatory process for identifying performance indicators
- Clarify the outcome. If the objectives of the intervention are clear and precise, indicator selection may be relatively easy. On the other hand, if the outcome explored is broad and unspecific, it is more difficult to define indicators.
- Identify possible indicators. Given that there are numerous potential indicators, it is worth identifying all possibilities
- Assess and select the best indicators. It is worth identifying the most valuable, appropriate and useful indicators. When choosing them, the aforementioned indicator selection criteria may be followed
- Fine tune the indicators

The data obtained in order to perform the evaluation must be individual or as disaggregated as possible, given that they may later be analysed as an aggregate or at the level of the programme, using multivariate statistical analysis or statistical inference. The data should not only encompass the individuals targeted by the intervention but also the control group in order to use counterfactual techniques.

[6] For example, in a programme that seeks to obtain outcomes in three educational centres, the chosen indicator is not plausible if it is at the national level.

Causal mechanisms in outcome evaluation: contribution and attribution

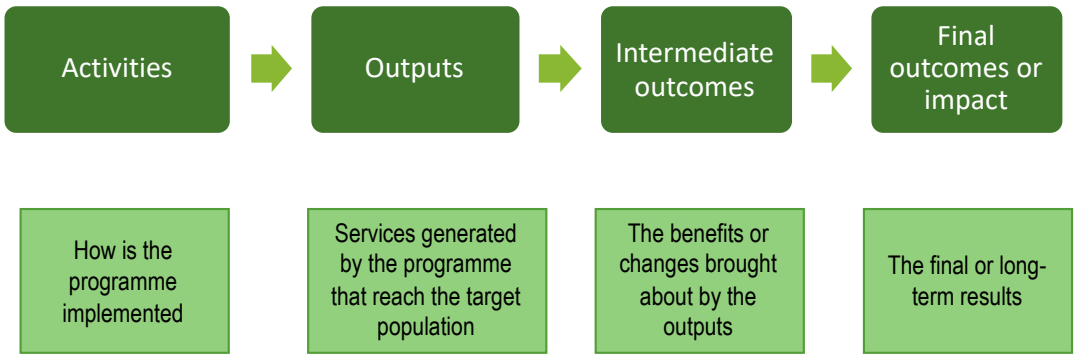
The section on concepts and definitions discussed two essential concepts in outcome evaluation: contribution and attribution. But, what methods or approaches may be used to unravel them? There are basically two evaluation methods, but regardless of which one is adopted, the analysis of the contribution of the programme to solving the problem and/or the degree to which the change or outcome may be attributed to the programme must be mentioned.

Causal theories are an essential part of evaluation, beyond the methods or tools employed. These theories must be based on a series of postulates: the creation of falsifiable theories, which are internally consistent, carefully selecting the dependent variables that define the phenomena to be explained, and maximising specificity (avoid abstract concepts or ones that cannot be empirically observed) (G. King, R.O. Keohane and S. Verba, 1994). Essentially, "good theories respond to sensible questions, they are causal, they may be verified with unobserved data, they are general and parsimonious, they are new and... they are not obvious" (G. Fontaine, 2015).

It is important to exhaustively and critically develop the entire chain of causality or causation, such that the association and the degree of links between the intervention and the outcome is indisputable, that is to say, there is the most complete or highest level of plausibility possible. Although this is possible in the case of initial or primary and intermediate outcomes, where the link with the outputs is easier to establish, it is more complex in the case of final outcomes and impacts, where other types of factors contribute to their achievement. Even changes at the intermediate level do not necessarily translate into satisfactory final outcomes.

An initial approach to the contribution/attribution of the intervention is that which uses methods or tools to determine the existence of a ***direct link between the intervention and the outcome***. Essentially, it deals with the results chain, the theory of change or the logic of the intervention. The underlying idea is to find plausible associations (Hendricks, 1996) that unequivocally support the conclusion that the observed outcomes or changes observed have been provoked by the identified cause or causes. In the case of an intervention, these constitute the deployed activities and actions, their outputs and outcomes. The results chain is a depiction of the hypothetical functioning of the programme and how it functions in practice, what are the outputs and the reasons behind their influence in the outcomes. It is a non-statistical comparative method that establishes patterns or standards that are consistent with the causal relationship based on the theory of change on which they are assessed by the evaluator.

Figure 6. Activities-results chain. Types of results. Source: Author's own based on J. Mayne (2001).



A second method for approaching attribution is the **counterfactual** method, based on establishing what would have occurred in the absence of the intervention. Counterfactual analysis is one of the most refined methods for determining attribution, a quantitative approach that compares the outcomes in the targets of the intervention (treatment group) and the outcomes in those that were not targeted (control group) in order to identify what part may be attributed exclusively to the programme. There are multiple forms of counterfactual analysis:

1. **Experimental methods**. The assignment to the treatment or control group is randomised and prior to the programme implementation, which is why they are also called randomised controlled trials. The advantage of these methods is that they tend to reduce or eliminate problems of selection and endogeneity bias, so that treatment and control groups are identical. Additionally, they allow us to capture baseline information, which has evident advantages when compared to the rest of the methods. However, they are problematic from the point of view of ethics and public policy implementation. Additionally, the evaluated object must fulfil a series of prior requirements: the intervention must be important, clearly defined and well-implemented, the evaluation must have sufficient resources, and preliminary evidence suggests that there are positive outcomes, although uncertain (GAO, 2009). These methods have traditionally been considered as the most rigorous, even when other methods make it possible to obtain or recreate very similar results (T. Cook, 2010).

2. **Quasi-experimental methods**. These methods are used when the intervention has been implemented or is completed. Assignation is not randomised, therefore it is necessary to monitor for selection bias. They are more commonly used in evaluation in developed countries, and when the evaluation is commissioned after the programme or policy design has been completed. Among the specific techniques that have traditionally been used are regression discontinuity design or difference in differences, as well as certain matching methods such as *Propensity Score Matching*[7], although there are other techniques that can perform statistically rigorous counterfactual analyses[8].

3. **Other methods**. Hypothetical counterfactual methods, causation models (linear or logistic regressions), instrumental variables, etc.

Counterfactual analyses of attribution through statistical "impact" techniques have a series of indisputable advantages. They focus on social phenomena, possess internal validity, are transparent, isolate the factors involved in the outcome and help to establish the degree to which the intervention provokes the change. At the same time it allows us to make predictions, make exact or very close measurements of social phenomena and make causal inferences. Their main disadvantage is that which is common to empiricism and positivism, and it is their inability to positively identify other causes that cannot be observed empirically, the dependence of indicators assumed to make an objective measurement of the reality but are only able to apprehend it partially.

"Impact" evaluation therefore, does not provide absolute certainty with regard to the outcome, but it provides certainty, which in turn depends on the design of the intervention itself, the availability of data, and knowledge of the external factors that affect the goals. The absence of all of these conditioning factors may lead to inexact measurements and explanations of public policies. The mechanisms of programme assignment must be acknowledged. In comparison to these techniques, the interpretive method attempts to limit uncertainty by means of inductive and deductive methods.

[7] Propensity Score Matching is one of the most widely used matching techniques (matching treatment group and control group by optimal scoring based on observed characteristics) in quasi-experimental methods.
[8] In recent years, econometric studies have provided other techniques of great interest, especially for overcoming certain practical limitations of the more commonly used techniques mentioned previously, depending on the availability of sources of information, microdata and conditions of the object of evaluation. It is not the aim of this guide to delve into their theoretical underpinnings, nor their practical applications, as there exists a sufficient body of work for this purpose. Apart from the abundant literature on econometrics available in English, two noteworthy publications in Spanish are (Gertler, Martínez, Premand, Rawlings, & Vermeersch, 2012) and (Perez López & Moral Arce, 2015).

Another advantage of these methods compared to experimental or quasi-experimental methods is that the latter are difficult to implement in interventions with total coverage, where it is impossible or nearly impossible to identify a control group that has the same characteristics as the population subjected to the treatment or intervention. “In most cases [of assessment of impact], it is difficult to attribute rigorously broad effects on different groups and at different levels over time to a specific intervention or set of interventions. Systems theory approaches typically provide more appropriate and useful tools for dealing with complex adaptive systems, e.g. Societies” (Ministry of Foreign Affairs of Denmark, 2006).

Contribution and attribution analyses constitute an ideal method for identifying the part of the change or the outcome that corresponds specifically to the intervention, as well as to identify the other factors that contribute to said change or influence the effects of the programmes. Both contribution and attribution approaches are recommended, along with existing methods of a more inductive nature, or causation methods.

Additionally, it is worth ascertaining attribution using, whenever resources are available, approaches that let us compare and display the existence of direct and unmistakable links between outputs, outcome and impacts, following the theory of change, as counterfactual analysis. Combining these approaches makes the outcome evaluation more rigorous.

Finally, it is important to highlight some recommendations that are perhaps not sufficiently mentioned in the guidelines or literature on outcome evaluation: the need to pay special attention to the context or process in which decisions are taken, and the programme’s environment; the need to include the analysis or evaluation of information requirements in the initial stages of the process in order to determine the process of causation (Davidson, 2000) -an aspect that is part of the considerations on the evaluability of the programme or policy- and finally, to ensure that the outcome evaluation is not exclusively focused on statistical adjustments in models, but that causal relationships must be established when comparing and eliminating alternative causal explanations, using quantitative as well as qualitative techniques.

Processes of outcome evaluation

There are two processes for outcome evaluation that generally reflect the two approaches and concepts of outcomes that have been explained in the first part of these guidelines.

1. **Prospective process**, which follows the results chain and analyses the generated outputs and how they are transformed into outcomes and impacts.

2. **Retrospective process**, where the starting point is the outcome. The outcome evaluation begins by revising the changes in the outcomes themselves, proceeds to analyse the relevant factors that influence them, and finally, determines the contribution of the intervention, the stakeholders and the context, to the outcome. That is to say, it is a process in reverse. The UNDP considers it to be the standard procedure for outcome evaluation (UNDP "Evaluation of results-based management at UNDP", 2008).

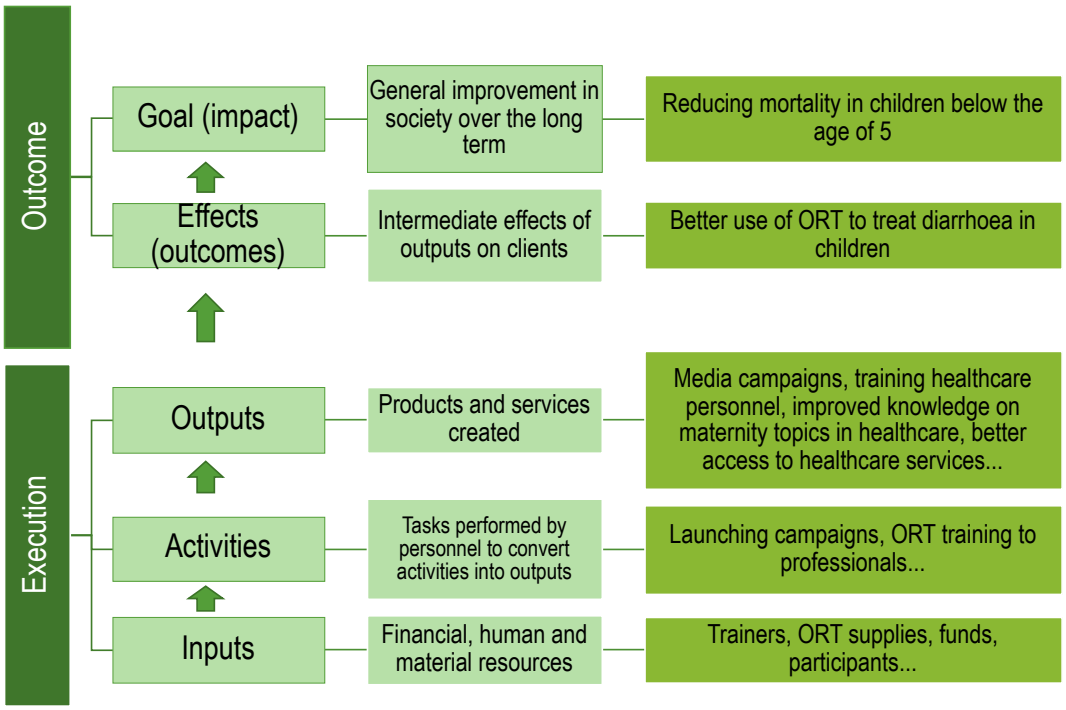
Both have a common element, the outcome; although it occupies a different position, it has features common in both, therefore it is important to establish certain guidelines on what is the outcome to be examined.

1.- Prospective process in outcome evaluation

Outcome evaluation adopts a sequential order focused on the intervention, essentially following the results chain or the theory of change. Thus, the inputs are activities which are then translated into outputs, that produce effects or (primary or secondary) outcomes and finally, tend to provoke a series of final outcomes or impacts. The point of reference is always the intervention, the internal processes or elements of the programme and how they are structured in order to produce the outcome and impacts.

This is the model followed by the World Bank. It may be exemplified in a programme from Binnendijk (2000). The objective was to improve infant mortality rates; media campaigns, training healthcare personnel, improving knowledge on maternity topics in healthcare and providing better access to healthcare services constituted the outputs, which were part of the execution and not the outcomes. The availability of these outputs led to an improved use of Oral Rehydration Therapy (ORT) in treating diarrhoea in children, thus partially fulfilling the final objective, which was to reduce mortality in children below the age of 5.

Figure 7. Execution-results chain. Source: World Bank, based on Binnedijk, (2000).



This sequence which ranges from the inputs to the outcome, must be supported by a theory of change or causal logic of the intervention which explains the implementation process until it arrives at the outcomes. An adequate intervention or programme must have a proven theory of change.

Although there are different tools that may be used to segment the process of transforming the inputs into outcomes (realist matrix, triple column, Tiny Tools, results chain, etc.) the most frequently used ones are the results chain and the logical framework matrix. Detailed information on how to use these tools is available in the Guidelines for the Design Evaluation of Public Policies and the Guidelines for the Implementation Evaluation of Public Policies. On general lines, the sequence to be followed in this approach is the following:

- The first step of the analysis is usually ***identifying the problem*** of the intervention. This identification is related to the suitability of the intervention and is not always present in outcome evaluation. From the comprehensive assessment approach to public policies, it is always worth analysing the complete life cycle, i.e., the design, implementation and outcome. However, if this comprehensive assessment approach is not followed, then it is necessary to at least outline, in varying degrees of depth, the characteristics of the problem and its causes and resulting effects. A detailed description of how to perform this analysis is included in the Guidelines for the Design Evaluation of Public Policies. One of the most noteworthy tools that may be used is the creation of problem trees.

- Secondly, the ***hierarchy of objectives*** must be established or validated. All public interventions must have a general objective that splits up sequentially into strategic and operational objectives and which are related to the measures that seek to achieve said general objective by means of a series of resources and activities.

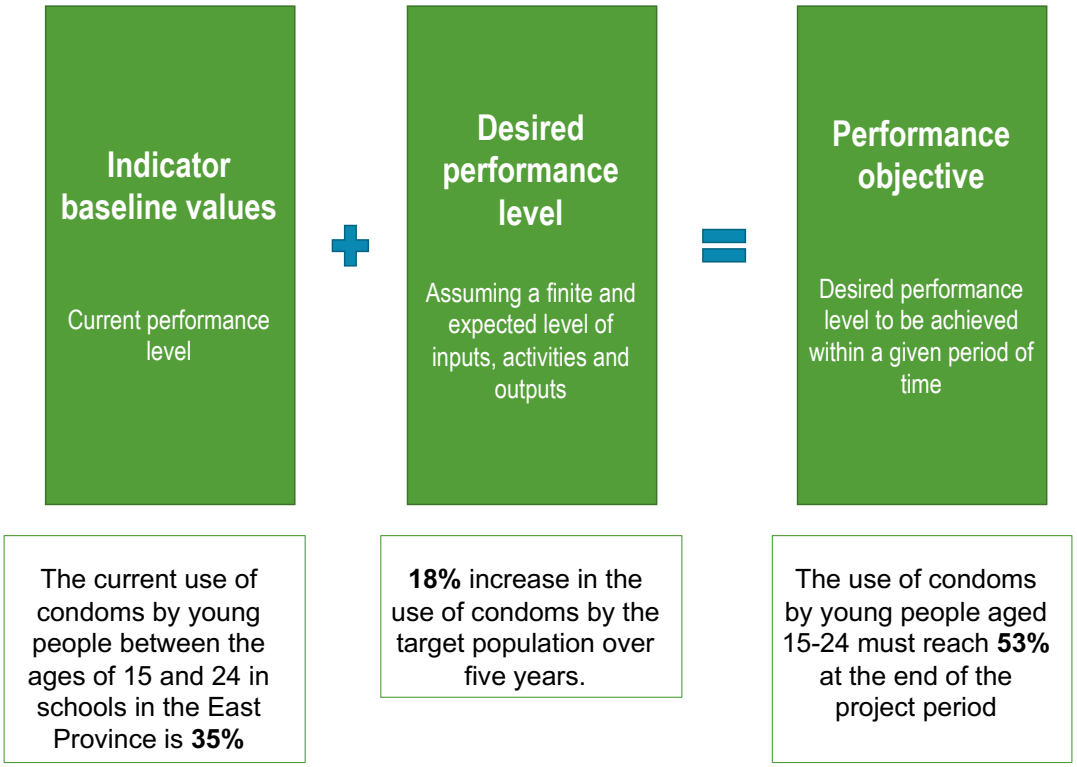
Nevertheless, the hierarchy of objectives is not always clear or explicit in the intervention, or it may be inverted at some level of the objectives. It is therefore necessary to build or validate it by means of different techniques with the participation of key actors. The importance of identifying the structure of objectives lies in its importance, when analysing the coherence of the design with the problem to be solved, with other interventions or with the measures or activities to be defined.

Evaluations based on objectives tend to analyse outcomes exclusively based on the targets, strategic or end objectives established by the programme or intervention, if these are explicit. Nevertheless, all evaluations must critically assess the consistency between the problem that is to be solved and the strategic or end objectives that have been established, given that occasionally, the objectives do not seek the manifested goal or do not correspond to the real problem to be solved.

Once the analytical hierarchy of objectives has been identified, the following step is to relate them to the activities, the measures designed to achieve them, with the allocated resources and with the offered products and services.

At this stage, the thresholds of the performance objectives to be achieved by the programme must be clearly identified. They may be mentioned explicitly by the intervention in its objectives, otherwise they must be identified by the evaluator(s). Similarly, the status of the situation with regard to the baseline objectives (before the implementation of the intervention) must be acknowledged. This will aid in understanding what is sought to be achieved, with regard to the initial status, given that the awaited objective performance is the actual performance level or status plus the desired level of improvement, as seen in the following figure.

Figure 9. Awaited objective performance and baseline. Source: Kusek and Rist, (2004) adapted by R. Rodriguez-Garcia and J. Zall Kusek (2004).



- Thirdly, the evaluator must obtain and analyse the information from ***the context in which the intervention takes place***, which conditions to a high degree the intervention and therefore, the outcome obtained.

- Establish the ***causal theory*** of the intervention. That is to say, a set of assumptions, conditioning factors and external factors that describe how and why the programme is meant to function and obtain certain awaited outcomes.

The theory of change is defined as how causal reflective reasoning explains the strategic options and outlines the premises of an intervention logic based on a desired change. Projects are drawn up and implemented within a logical framework to make the strategic decisions defined in the theory of change, operational. All public interventions must be based on a causal theory. It refers to how the intervention seeks to generate, and generates, the required changes at different stages or phases to achieve the intermediate outcomes and the expected final transformation. It defines the strategies to be followed, taking into consideration the risks and factors that may influence the achievement of these changes, as well as the conditioning factors that lead to their achievement.

The logic is the following: using a series of inputs (human and financial resources, etc.) a series of activities are performed that provide outcomes regarding the operational

objectives. These in turn lead to changes in the affected variables (strategic objectives) and the socio-economic environment of the intervention (general objective).

The use of the results chain and the logical framework matrix in outcome evaluation presents certain common and specific aspects of their use in the design of public interventions and their evaluation, or in their management or implementation.

The analysis of the theory of change on the design of the intervention has a two-fold goal: A common element of both tools is that they provide a better understanding of the initially projected intervention and correctly arrange the different components of the programme and the relationships between them. It allows us to assess the consistency and quality of the underlying theory of the intervention at the level of the defined strategies, i.e., how to solve the requirements or problems of the target population, and the contribution of the implemented measures or activities with regard to the distribution of the input resources, and outputs or services aimed at solving the problem. However, an analysis must be performed a priori on whether the underlying logic was consistent, adequate and permitted an objective achievement of the outcome. Additionally, it permits or guides the formulation of the evaluation questions.

- Analyse the ***deployment of the intervention***. Even if an in-depth evaluation of the process or implementation is not performed, the inputs used, the activities or processes implemented, and the services or outputs generated must at least be mentioned, as they precede the outcomes and influence their achievement. If there are defects in the implementation or distortions are detected, it may be questioned whether the objectives can be achieved.

- Analyse the ***degree of achievement of the objectives or the outcome or changes produced***. This stage has elements in common with the retrospective process which is discussed later (determining the status of the outcome). Special attention must be paid here to:

  - The outcome may not be identical for all the targets of the programme or policy, given that there are initial and specific factors or situations that affect the outcome. For some targets, there may not even be any effect. The evaluation must establish the gradation or breakdown of the different effects or outcomes produced.

  - The analysis of the factors that have influenced the outcome either in isolation or jointly with the policy or programme (contribution analysis), as well as, when possible, the exclusive attribution of the evaluated intervention. Here, the steps mentioned in these Guidelines and in the retrospective process may be followed.

- Compare the formal or objective theory on which the programme is based with the practical reality: whether the theory of change has functioned correctly, if its bases have been fulfilled, if the activities were performed in the implementation and the stipulated outputs have been generated, or if there have been deviations. In short, whether, the outcomes produced are suited to the theory or not, and what are the factors behind a mismatch: an inadequate design, an incomplete deployment, the influence of contextual or external factors, etc.

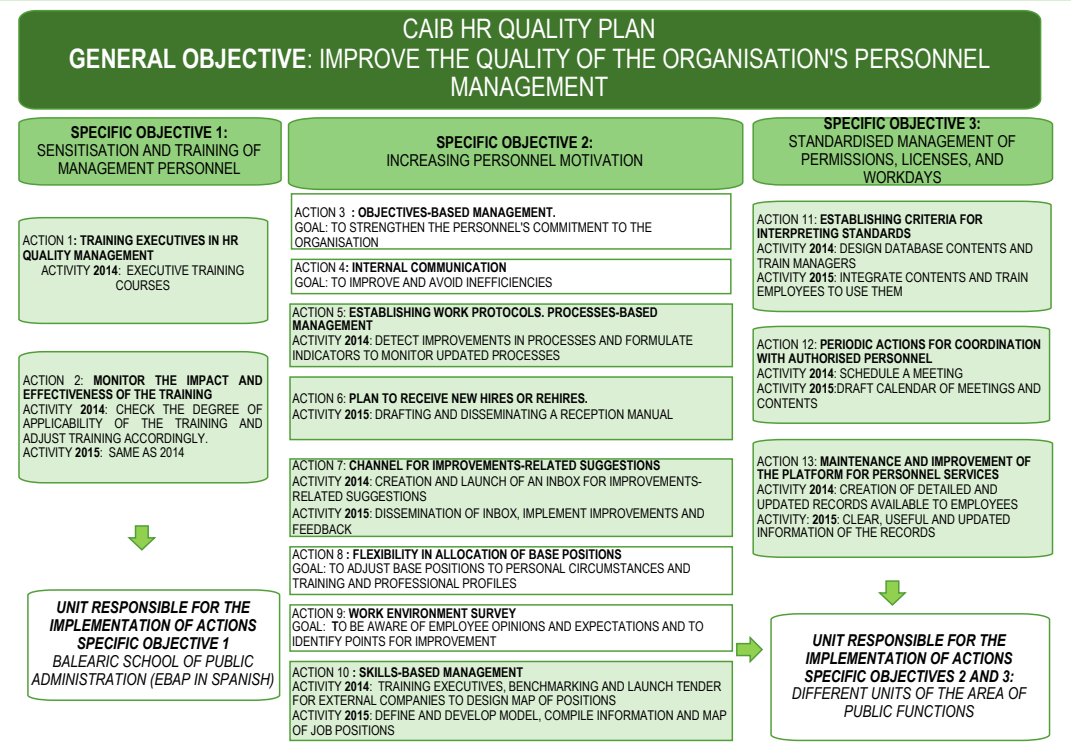
Objective tree

One of the most frequently-used tools to graphically arrange these analyses is an ***objective tree***. The objective tree is a methodological procedure that identifies and classifies objectives according to their importance and displays the means-goals relationships in a diagram. "It consists of converting the negative states of the problem tree into solutions, expressed as positive states (...). As a matter of fact, all of these positive states are objectives and are presented in a diagram of objectives which displays the hierarchy of the means and the goals" (CEPAL, 2005) Here, the core problem would be the main objective and the effects would become the goals of the intervention. The objective tree is built in the following manner:

- The problem is converted into a desirable positive state.
- Analyse the causes of the problem, formulate them as positive states and adjust them to the general objective.
- Analyse the means-objectives-goals relationships to ensure the consistency of analysis.

For example, in the "Evaluation of the Human Resources Quality Plan of the general services of the Administration of the Autonomous Regions of the Balearic Islands" (AEVAL, 2015), the general objective was to improve the quality of the organisation's personnel management. To achieve this, three specific objectives were proposed: the sensitisation and training of management personnel, increasing personnel motivation and the standardised management of permissions, licenses, and workdays. These objectives in turn break down into different actions that are the responsibility of different units.

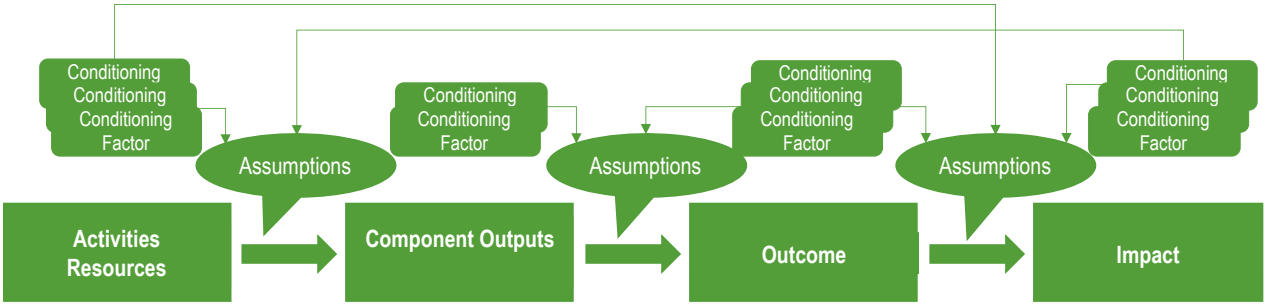
Figure. Objective tree. Source: Evaluation of the Human Resources Quality Plan of the general services of the Administration of the Autonomous Regions of the Balearic Islands (AEVAL, 2015).



Results chain

The results chain consists of a linear representation with boxes/ squares of objectives, activities, products, outcomes and impacts. It is a useful tool for representing simple interventions. For more complex interventions, the logical framework matrix is more suitable than the results chain.

Figure 10. Diagram of the theory of change as a results chain. Source: Author's own.

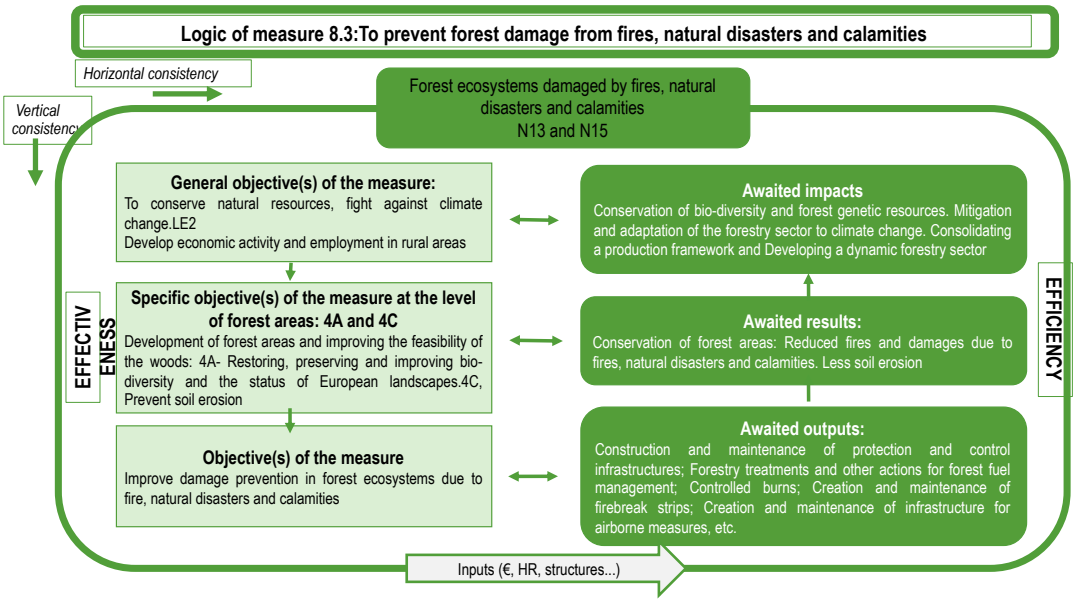


Logical framework model

The ***logical framework model*** is a matrix representation that provides a horizontal and vertical reading to obtain the causal relationships and the narrative of the logic. This technique has limitations, as it does not reflect all the interactions. The narrative derived from the logical framework matrix is the basis for the theory of change, which must be supplemented with the analysis of the assumptions and conditioning elements (internal and external factors to be taken into account for achieving successive assumptions and finally for achieving the vision).

On the basis of the logical framework matrix, we may define the timeline of the activities and the indicators of each for their monitoring. It is not only a planning or management instrument, but is also used in outcome evaluation.

Figure 11. Logical framework matrix. Source: Intermediate Evaluation of the National Plan for Rural Development (AEVAL, 2017).



2.- Retrospective process in outcome evaluation

Generally, the sequence to be followed in this approach, as explained below, follows the UNDP model, but it differs with regard to the content of each component:

- **Determining the status of the outcome**. The first step towards a good outcome evaluation is to check the status of the outcome, which constitutes the starting point. "The network of analysis tends to broadly cover all that has been undertaken within the scope of the intervention and beyond, given that they may be deemed to influence the outcome in question". The steps to be followed to determine the status of the outcome are the following:

  - Firstly, the evaluator must *determine the outcome sought to be achieved* and the changes required to solve the problem that is the reason behind the intervention. Initially, the strategic outcomes of the intervention that are awaited or desired may be taken as reference, which may be ascertained from the documents on strategy, management, essential regulations or declarations. These outcomes which are objectively considered will function in the evaluation as standards against which to compare the obtained outcome or the changes that are really produced.

  - Secondly, it is important to *identify or establish what is the current state*, objectively considered, prior to the intervention and from the point of view of the outcome (obtaining baseline information). This is done by means of suitable indicators or references, or the management system, if it exists.

  - Thirdly, the evaluator must obtain and analyse *information on the context* of the intervention.

  - Fourthly, *the objective status of the produced outcome or change* must be established, within the terms and considerations previously mentioned in these Guidelines. For this, all measurement tools that allows us to understand the produced change or outcome, of any type, may be used. The indicators for measuring outcomes or achievements must be exhaustive and must permit the measurement of change in all aspects and dimensions. Other tools to establish the status of the outcome are interviews of stakeholders, managers or experts involved, in order to acquire their opinion of the change or outcome produced, qualitative or quantitative techniques, target population surveys, etc.

- **Factors that affect the outcome**. A second step is to perform an exhaustive analysis of all the factors that affect the achieved outcome. The analysis of the factors is, along with the outcomes or impacts, a central aspect of outcome evaluation. Only on very few occasions is the change the exclusive outcome of a public intervention, given that many factors and interventions are frequently involved within the same issue in highly complex contexts. Additionally, contextual factors themselves may help or hinder the achievement of outcomes or condition them. They may also be inter-related or influence each other. The factors are a test of the intervention hypothesis and its underlying assumptions (UNDP, 2008), given that an intervention that does not constitute a direct and fundamental factor of change is hard to justify.

  The steps or the process to be followed at this stage are the following:

  - The factors that contribute to the produced change must be identified, and their extent or importance with regard to the change must be determined, along with the direction of their influence. The key factors must be analysed in depth. Similarly, it is important to analyse the interactions that take place between the factors.

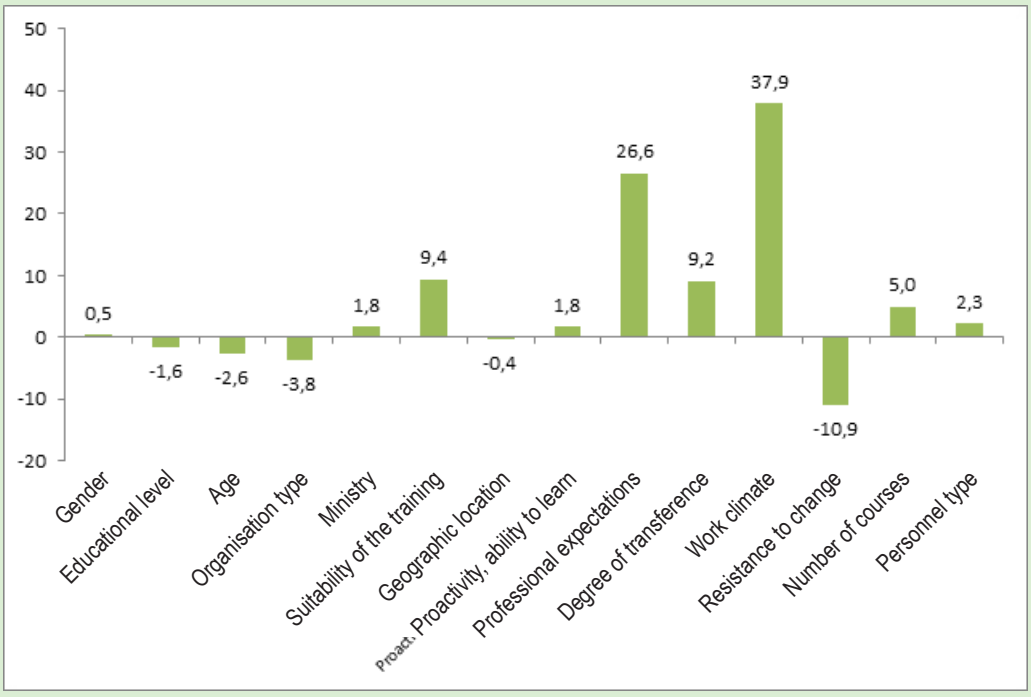
  - Additionally, the factors that are responsible for the different degree of achievement of the outcomes must be identified, which may vary, depending on the participants or the targets. The circumstances responsible for the satisfactory or unsatisfactory outcomes must be acknowledged. Some of the causes that may justify or influence the outcomes include demographic and social conditions, levels of difficulty or varying needs of the target individuals, etc., which should be studied.

  - Context analysis is essential in any evaluation, including the initial context as well as the context after the change has taken place, as it may shed sufficient light on the responsible factors. The intervention itself may not have been conceived as the sole factor that leads to the change but it is responsible for changing or modifying other contextual factors. Changes must be understood in a dynamic manner and interventions usually bring about multiple changes in the environments, even when they do not produce outcomes.

  - The unintentional effects or the motives and factors that play a role in the inability to achieve the objectives, targets or change must be analysed and defined.

Example of identification of factors or phenomena that affect the outcome. Evaluation of the training of public employees of the General State Administration Services (AEVAL, 2013).

The end objective of training government employees is to improve their professional performance, by the acquisition of necessary skills and knowledge, and their transfer to the position in question. This evaluation identified and analysed the factors that contributed to the improved professional performance of government employees, and not just those related to their training. These factors were previously identified by means of a factorial analysis and they consisted of: personal factors; factors linked to training received (satisfaction with the training and its design and application), factors related to the professional environment; factors related to working conditions. Multiple linear regressions were performed, of the factors identified by the factorial analysis, as well as personal variables or phenomena, variables relating the work environment, the organisation, etc. The following figure displays the conditioning elements of the variation in professional performance experienced in 2013-2014. Regression model-beta coefficients



Thus for example, the report states that "(...) the second aspect that determines professional performance are the professional expectations perceived in the workplace. It is an aspect clearly linked to the institutional and organisational context. When employees deem their chances of improved career prospects to be scarce or null, their performance does not improve. On the other hand, the greater their expectation of improved prospects, the greater the perceived improvement in their performance".

- **Determining the contribution and attribution of the programme to the outcome**. The causal nexus between the intervention, its outputs and effects, and the change must be clearly established, given that the credibility and rigour of an evaluation are largely dependent on it: "Credibility is highest when outcome evaluations are perceived as indicative of rather than inventing links between the outputs and the outcomes." (UNDP, 2008)

  Here the use of counterfactual methods becomes especially important, provided the microdata and variables needed for the analysis are available. Thus it performs a rigorous identification of the degree to which the change or outcome is provoked solely and exclusively by the intervention.

Example of identifying attribution using counterfactual methods. Evaluation of the training of public employees of the General State Administration Services (AEVAL, 2013).

In this evaluation, once the different factors or characteristics that determined improved professional performance were identified, a counterfactual analysis was performed to identify the part of the improved professional performance that was due solely and exclusively to the training received by the government employees of the General State Administration Services. This was achieved by means of multiple Propensity Score Matchings that measured different degrees of intensity of the training received. Different matching algorithms were used to build the contrast. The results of the technique were checked with difference in differences. Results demonstrated that with higher training intensity, professional performance improved by 5-8%, generated by ATT (Average Treatment effect on the Treated).

Report on the Evaluation of Rebates for Hiring Persons with Disabilities (AEVAL, 2009)

To analyse the impact of the measure to be evaluated (rebates on Social Security contributions in permanent contracts for people with disabilities), different Propensity Score Matchings were performed using data from the Continuous Work History Sample, as well as contract records. The outcome variable was job permanence, either the continuation of the same contract or other contracts that the individual may have signed (thus, two target variables). Permanent contracts for persons with disabilities were compared with 5 contract modes: ordinary permanent contracts, permanent FEE (Promoting Stable Employment) contracts, conversions of permanent FEE contracts, ordinary conversions, and conversions of disability contracts. The Propensity was replicated with different variables of interest, such as gender and company type.

Other statistical inference tools (regression, for example) were also useful for establishing the size of the programme or intervention's effect. One of the benefits of regressions is being able to capture the outcome of the intervention along with a series of potential explanatory factors that are responsible for the outcomes. That is to say, they make it possible to analyse the contribution of other factors as well as the attribution of the intervention, as may be seen in the following example.

Example of identifying contribution and attribution (jointly) using linear regression. Evaluation of the measures for streamlining and improving the management of Temporary Disability (AEVAL, 2009).

This evaluation uses multiple linear regressions, on one hand, to determine the variables or phenomena that condition the duration of the processes of leave for common contingencies, and on the other hand, the extent to which some of the measures considered influenced said duration. Using the Continuous Work History Sample which contains numerous variables on work and personal characteristics (type of contract, Social Security system, contribution group, temporary disability management agency, size and sector of the employing company, age, etc.), regressions were performed including two specific measures: the attribution of the management of leaves longer than 18 months to the National Social Security Institute (INSS in Spanish) and the existence of pilot programmes for monitoring disabilities. The results showed that the first measure had no effect whereas the second did contribute to reducing leave duration to a certain extent.

EVALUATION CRITERIA AND QUESTIONS

Evaluation criteria in outcome evaluation

According to the Institute for the Evaluation of Public Policies and this Guide, evaluation criteria are the different points of view or approaches to the target of the evaluation, based on evidence, for its assessment. They are conditions, rules and also "principles, standards or ideas on assessment, based on which the evaluated object is assessed" (García Sánchez, 2010).

The criteria act as guidelines to formulate questions and focus the evaluation, giving it a structure that "covers the field or dimensions of a public policy or programme to be evaluated. Indeed, the questions themselves usually belong to different sets of criteria" (AEVAL, 2015).

There are numerous criteria to be considered in outcome evaluation. The design (internal coherence of the programme) evidently conditions the outcome, as does the suitability or the correct identification of the requirements and the public problem, the implementation, the coverage of the programmes, and their external coherence or complementarity. Given that these have been studied in the other guidelines, we shall not delve into all the criteria but only those that may be referred to in outcome evaluation, such as effectiveness, efficiency and sustainability.

Effectiveness

Effectiveness has a double dimension. On one hand, it focuses on the fulfilment of the objectives, mainly strategic ones of the public programme or policy. The objectives must be related to the awaited outcome of a programme, either explicit or, on the contrary, derived from a correct identification of the objectives. On the other hand, effectiveness is related to the extent to which the problem that motivated the intervention has been solved or mitigated. Another approach to effectiveness or achievement is that which focuses on the degree of change in the reality or phenomenon, and which explores if there have been positive or negative changes, either intentional or unintentional, as an outcome of the intervention. Once these aspects have been determined, it is necessary to establish the degree, form, and intensity to which the changes, the fulfilment of objectives, or the resolution of the problem are due to the intervention.

Efficiency

The extent to which the desired outcome and/or effects have been obtained with the least amount of required resources. Efficiency evaluations are a specific type of evaluation that go beyond outcome evaluation. As a matter of fact, some authors ((Rossi, Lipsey and Freeman, 2003) place it on a higher level within the hierarchy of evaluations. In practice, "effectiveness and efficiency" evaluations are highly sought after.

Sustainability

This criterion allows us to assess whether the implemented measures and their outputs may be sustained over time in order to maintain or boost the outcome, or if it is necessary to retool them based on the context, resources or other factors that may have changed or are projected to change.

Evaluation questions

The system used to create "the logical structure of the evaluation" based on a series of "queries and hypotheses that will make it possible to implement the evaluation" (AEVAL, 2010), consists of the evaluation questions, the basic unit of research in an evaluation. This logical structure includes the scope and focus of the evaluation and therefore, steer the evaluation design.

The evaluation questions constitute the operationalisation of the evaluation criteria, are framed on the basis of said criteria, and can be broken down into questions and sub-questions.

Evaluation matrix

The ***evaluation matrix*** is the basic tool that steers the evaluation process. Normally, it consists of the evaluation questions and criteria and usually includes the indicators, sources, tools and techniques of analysis. It seeks to make robust and credible conclusions and recommendations that are generated from the responses to said questions, according to the involved criteria.

The following section shows an example of an evaluation matrix with the common criteria for an outcome evaluation[9].

[9] The evaluation questions that are listed in the matrix are not exhaustive.

EVALUATION MATRIX

I. To what extent has the problem, demand or requirement responsible for the intervention been solved?				
SUB-QUESTIONS	CRITERIA	INDICATORS	SOURCES	TOOLS
1.1. Have the objectives established in the intervention been achieved, with regard to the awaited outputs of a programme?	**EFFECTIVENESS**	% of fulfilment of strategic and operational objectives Analysis of the underlying causes of the failure to achieve the objectives	Monitoring reports Tables and checklists for monitoring Design, implementation and outcome documents of the programme Programme databases Variables, according to the target of the evaluation and the sector of intervention Interviews of stakeholders (managers, planners, experts and potential targets of the intervention) Target survey	Documentary analysis Semi-structured interviews of experts, managers, persons or bodies in charge and targets of the intervention Questionnaire for programme targets Descriptive data analysis
1.2. What immediate changes have been produced by the intervention in potential beneficiaries or the target population of the intervention?	**EFFECTIVENESS**	Analysis of the changes produced in the potential beneficiaries of the programme Quantitative indicators of the produced changes, based on critical variables and variables of interest Underlying causes or factors of the produced changes	Results and outcome reports of the Intervention design, implementation and outcomes of the programme Programme databases Variables, according to the target of the evaluation and the sector of intervention Interviews of stakeholders (managers, planners, experts and potential targets of the intervention) Target survey	Semi-structured interviews of experts, managers, persons or bodies in charge and targets of the intervention Questionnaire to potential targets of the programme and the control group (non-targets) Descriptive statistical analysis Linear, logistic regressions, etc. of dependent variables. Time series

SUB-QUESTIONS	CRITERIA	INDICATORS	SOURCES	TOOLS
1.3. What changes of effects have been produced in society in general, beyond the direct targets of the intervention? (effects sought and unsought, positive and negative)	**EFFECTIVENESS**	Analysis of all types of changes produced in society within the scope of the public intervention Quantitative indicators of all types of changes produced, based on critical variables and variables of interest Underlying causes or factors of the produced changes Variables, according to the target of the evaluation and the sector of intervention	Results and outcome reports of the intervention design, implementation and outcomes of the programme Programme databases Variables, according to the target of the evaluation and the sector of intervention Interviews of stakeholders (managers, planners, experts and potential targets of the intervention) Target survey	Semi-structured interviews of experts, managers, persons or bodies in charge and targets of the intervention Questionnaire to society and the potential targets of the programme Descriptive statistical analysis Linear, logistic regressions, etc. of dependent variables. Time series
1.4. To what degree does the programme or policy contribute to the observed results? Or: Can the intervention, by itself, have a significant weight or influence on the outcome?	**EFFECTIVENESS**	Analysis of the contribution of the programme or policy to the changes or to the achievement of the objectives Quantitative indicators of the degree of contribution of the intervention by the intervening factors in the change	Outcome and impact reports of the intervention Design, implementation and outcome documents of the programme Programme databases Variables, according to the target of the evaluation and the sector of intervention Interviews of stakeholders (managers, planners, experts and potential targets of the intervention) Target survey	Semi-structured interviews of experts, managers, persons or bodies in charge and targets of the intervention Questionnaire to potential targets of the programme and the control group (non-targets) Descriptive statistical analysis and descriptive techniques or classification or segmentation techniques Factorial analysis Linear, logistic regressions, etc. of dependent variables. Time series Other inferential techniques

SUB-QUESTIONS	CRITERIA	INDICATORS	SOURCES	TOOLS
1.5. What factors influence in the observed change or outcomes?	**EFFECTIVENESS**	Analysis of factors involved in the changes Produced Quantitative indicator variables according to the object of evaluation	Results and outcome reports of the intervention design, implementation and outcomes of the programme Programme databases Variables, according to the target of the evaluation and the sector of intervention Interviews of stakeholders (managers, planners, experts and potential targets of the intervention) Target survey	Semi-structured interviews of experts, managers, persons or bodies in charge and targets of the intervention Questionnaire to potential targets of the programme and the control group (non-targets) Descriptive statistical analysis and descriptive techniques or classification or segmentation techniques Factorial analysis Linear, logistic regressions, etc. of dependent variables. Time series Other inferential techniques
1.6. What part of the change or result may be solely and exclusively attributed to the public intervention?	**EFFECTIVENESS**	Indicators on attribution analysis: ATT, differences, etc. Other indicators according to the object of evaluation and the intervention sector	Programme databases Variables, according to the target of the evaluation and the sector of intervention Interviews of stakeholders (managers, planners, experts and potential targets of the intervention) Target survey	Semi-structured interviews of experts, managers, persons or bodies in charge and targets of the intervention Questionnaire to potential targets of the programme and the control group (non-targets) Impact evaluation techniques: psm, difference in differences, regression discontinuity, etc. Linear, logistic regressions, etc. of dependent variables. Time series Other inferential techniques

SUB-QUESTIONS	CRITERIA	INDICATORS	SOURCES	TOOLS
1.7. What improvements have been made to the welfare of persons or society as a whole, thanks to the intervention?	**EFFECTIVENESS**	Analysis of the improvements or shortcomings with regard to the welfare of the general population within the scope of the intervention Other indicators according to the object of evaluation	Programme databases Variables, according to the target of the evaluation and the sector of intervention Interviews of stakeholders (managers, planners, experts and potential targets of the intervention) Target survey	Semi-structured interviews of experts, managers, persons or bodies in charge and targets of the intervention Questionnaire to society Impact evaluation techniques Linear, logistic regressions, etc. of dependent variables. Time series Other inferential techniques
2.1. What characteristics of the obtained outcome condition their permanence over time?	**EFFECTIVENESS**	Logical-rational sustainability analysis of the outcomes based on: their specific characteristics, their intensity or degree, links to the outputs of the intervention in question, factors that contribute to their achievement, etc. Analysis of the necessary and sufficient conditions for their continuation over time Dependence of the results on factors related to: financial-budgetary resources, political-administrative context; structure of the managing organisations; the inputs-activities/processes chain-outcomes chain; necessary conditions within the scope of the intervention that promote or restrict the outcomes or performance; positions of the stakeholders and their expectations and interests	Evidence of the outcome evaluation, its characteristics, nature, breakdown, logical-causal chain, etc. Management reports or documents on inputs, processes, activities, financial-budgetary elements Analysis of the organisations Interviews of stakeholders (managers, planners, experts and potential targets of the intervention) Target survey	Semi-structured interviews of stakeholders (managers, planners, experts and potential targets of the intervention) Target survey Documentary analysis Semi-structured interviews of experts, managers, persons or bodies in charge and targets of the intervention Target survey Quantitative outcome and impact techniques

SUB-QUESTIONS	CRITERIA	INDICATORS	SOURCES	TOOLS
2.2 What are the risks of any type that may condition the permanence o the results?	**EFFECTIVENESS**	Analysis of risks and barriers and promoters of the outcome	Interviews of stakeholders (managers, planners, experts and potential targets of the intervention) and experts Target survey Documentary analysis Semi-structured interviews of experts, managers, persons or bodies in charge and targets of the intervention Target survey	Documentary analysis Semi-structured interviews of stakeholders (managers, planners, experts and potential targets of the intervention) and experts Delphi Method Nominal Group Technique Specific techniques for risk identification and evaluation

ANALYSIS TECHNIQUES IN OUTCOME EVALUATION

In order to perform an outcome evaluation, there are different tools and techniques that allow the evaluator to obtain rigorous proof that responds to the evaluation questions or to analyse the different questions mentioned in the evaluation.

These Guidelines provide a brief description of the most relevant social research techniques that are of the greatest use and validity for evaluation in general. The most traditional classification of available techniques is that which distinguishes between qualitative and quantitative techniques.

Thus, among ***qualitative*** techniques, we have documentary analysis, interviews, discussion groups, nominal group techniques, discourse analysis, SWOT analysis and case studies. And among ***quantitative*** techniques we have purely descriptive statistical methods: statistical inference or relations between the variables or phenomenon under study, either by means of statistical association or more complex analyses, such as simple linear regression models, multiple linear regression models, logistic regressions, etc. We also have impact evaluation techniques (*Propensity Score Matching*, difference in differences, regression discontinuity, etc.).

When analysing the outcomes, qualitative methods allow us to obtain in-depth information on the perceptions and opinions of a group of persons on a certain question.

On the other hand, quantitative methods allow us to respond to different questions: the external validity or degree of generalisation of the achievements or outcomes, the factors or causes behind a certain phenomenon or change, performance, input and results indicators, degree of implementation of the measures or resources employed.

Figure 12. Analysis techniques in an evaluation. Source: Author's own.

Type of technique		Purpose/nature
Qualitative techniques	Interviews	Exploratory. Applicable at any stage.
	Group discussion	Collecting qualitative information. Facilitating comprehension, credibility and acceptance.
	Nominal Group Technique (NGT)	Structured analysis of ideas and problems.
	Discourse analysis	Analysing all discourses and the contexts in which they are produced.
	SWOT	Reducing uncertainty and define strategies.
	Case study	Analysis of results and impacts.
Quantitative techniques	Survey	Obtaining descriptive information or other type of information in order to apply other techniques.
	Linear regression	Analysis of explanatory causes and estimating effects.
	Logistic or probabilistic regressions	Analysis of explanatory causes and estimating effects.
	Cost-benefit analysis	Knowledge of differentiated impacts. Efficiency analysis.
	Cost-effectiveness analysis	Effectiveness analysis based on a relevant criterion.
	ARIMA Models	Time-series analysis.
	Multi-level analysis	Studying contextual factors, either by hierarchy or by levels.
	Stochastic frontier models	Measuring efficiency in terms of input maximisation.
	Factorial analysis	Reducing underlying dimensions.
	Impact evaluation methods	Measuring net effects attributable to a public intervention.
Mixed Methods	Multiple criteria analysis	Structuring and combining assessments taken into account in a decision.

Quantitative techniques

Survey

A survey is one of the most frequently-used techniques in any type of evaluation, including outcome evaluation as it allows us to clearly identify questions related to design and the achievement of objectives or the results obtained, as well as the perception of the stakeholders and the existing difficulties from the point of view of the managers, stakeholders or the targets of the intervention. It also allows us to obtain results from a specific territory that may be generalised to the entire population. As a source of primary data, it allows the evaluator to arrange them in the most convenient way possible to obtain the necessary information for the research.

It is a reliable but expensive technique and requires an exhaustive knowledge of the intervention and a thorough preparation of the framework of analysis by the evaluator.

When performing a survey, the first step is the ***sample selection***, which must be as representative as possible of the reference population, in order to make generalisations with regard to the population. Random sampling methods ensure the best sample representation. This means that any individual in the selected sample has the same probability of being selected.

Another aspect to be taken into account to optimise the results of the survey is the selection of the sample size. This requires a considerable knowledge of sampling techniques, a topic which is beyond the scope of these Guidelines. Nevertheless, it must be remembered that the greater the sample size, the lower the estimation error and thus, the more significant the results.

On other occasions, when the total population is not excessively high, all the members may be surveyed. Let us take, for example, a survey of organisations or units numbering between 100 and 200.

Once the sample size is selected, we come to the ***survey design***, which is the instrument for compiling and measuring data, and is characterised by a series of questions arranged according to a specific logic. Its design must be adjusted to the established objective and for this, we must be clear about what we wish to ask and above all, how to ask: It is important for the questions to be clear and concise, and flexible and "comfortable" answers must be provided to the interviewee. Finally, the questionnaire must not be very long. There are different types of questions: open, closed, semi-open (or semi-closed).

With regard to the ***mode of administration*** of the questionnaires, they may be self-administered, in-person, telephone, postal or online surveys. The decision to opt for one or another depends on the advantages and disadvantages of each of them according to the topic under study, the available time and financial resources and the target population of the survey.

- In-person surveys are most frequently used in social research. They have the advantage of a more complete form of obtaining information and allow researchers to capture the environment surrounding the survey. But it has the disadvantage of being expensive, slow and difficult to access by certain populations.

- The main requirement for telephone surveys is that the surveyor must have a comfortable format. When drafting the questionnaire it is important to assess whether the design, duration, order, and interpretation are the most suitable. Currently they are mostly performed as Computer Assisted Telephone Interviewing (CATI), which lowers costs and the time required to perform them. However, it is not appropriate for delicate topics or complex questions. This survey mode may suffer from technical errors.

- In a postal or online survey, the interviewee reads the questionnaire and notes down their responses. There is no interviewer and therefore, a letter of presentation is required. It is a cost-effective technique and requires few personnel to perform the survey. It gives anonymity and flexibility of time to the interviewee. Its disadvantages include low levels of response and errors in filling out the questionnaire.

With regard to ***specific types of surveys***, we may mention:

- Omnibus surveys that allow us to include various topics, research or evaluation goals in a single survey. It is cost-effective, as instead of multiple surveys, only one is performed, thus sharing the research costs, and formulating a reduced number of questions in the same questionnaire and targeting the same sample. This type of survey is generally meant for large populations to achieve a financially feasible study. The questionnaire follows the same criteria as the interview but distinguishes itself by being arranged into different sub-questionnaires or modules with regard to different topics or outputs.

- Panel survey is a quantitative market research technique that is performed periodically on the same representative sample of a specific population.

Descriptive statistics

Once the survey is performed and the data has been filtered, they are analysed by means of ***descriptive statistical techniques.***

- Absolute and relative frequencies (the number of times an event is repeated and what it represents at the level of the population, respectively).

- Measures of centralisation are used (mean, median and mode) to obtain an overview of the data.

- Measures of dispersion, that provide an idea of variation in the sample data. They are useful when assessing the reliability of measures of centralisation such as the mean. They have an inverse relationship, the higher the measure of dispersion, the lower the representativeness of the measure of centralisation. The most well-known are variance and range; the range measures the difference between the maximum and minimum value that the observations can reach; variance measures the distance between the data and the mean.

These descriptive statistical techniques are characterised by their study of random phenomena; therefore their results are not precise and are accompanied by a certain degree of uncertainty. To measure this degree of uncertainty, we use statistical inference techniques.

Statistical inference

Statistical inference techniques give us the answers to questions such as: What variables influence the incident? How do the variables influence the incident? Is it possible to obtain a model that explains the incident and allows us to predict its behaviour? Some of these techniques are described below.

Impact evaluation methods

Impact analysis or evaluation methods allow us to determine what part of the observed effects or results of a phenomenon are solely and exclusively attributable to a fact, in this case, a programme or an intervention. They are also called counterfactual methods.
The advantage of these methods is that they statistically isolate multicausality and isolate the effects, so that we may state with statistical rigour that the observed results are the result of a factor, fact, programme, or intervention.

Impact evaluation methods compare the results observed in the population, drawing a distinction between the target group or persons who receive an intervention and those who do not receive it, called the control group. If both groups are statistically similar or identical, the observed result can only be dependent on the treatment.

Impact evaluation tools or techniques may be divided into those based on experimental models, when it is possible to define in advance the phenomenon that receives or does not receive the intervention, through random processes; and quasi-experimental models where it is not possible to randomise in advance. Of the latter, noteworthy methods are difference in differences, Propensity Score Matching and regression discontinuity. It is not the goal of these Guidelines to describe these methods in detail, especially since there is a wealth of technical literature on them.

Linear regression

In didactic terms, regressions seek to explain a variable or phenomenon that is deemed independent or endogenous by means of a series of facts, phenomena or variables that are called regressors, covariates or explanatory factors. It is the latter that may explain to a certain degree a phenomenon, behaviour, or reality.

Regression allows us to adjust a point cloud to a function where the endogenous or independent variable is explained partially through regressors or dependent variables, at the same time that the contribution of each dependent variable to the aforementioned explanation is determined. The difference between the real values and the explanation of the endogenous variable by the regressors is what constitutes the error term or random term.

When the independent variable is continuous and the function that links the endogenous variable with the regressors is linear, it is called linear regression. Apart from this configuration element, the assumptions on which the adjustment is made are: non-correlated regressors, their variance is constant (homoscedasticity), the errors in the measurement of each are inter-related and add to the total error, and the expected value is equal to zero, that is to say, the errors of a similar magnitude and opposite signs are equiprobable.

Provided the target of the analysis permits it and there is sufficient high-quality data, this technique can provide useful evidence for an evaluation.

Example: Evaluation of the Plan for Measures to Improve Cross-Border Healhcare Services (AEVAL, 2013).

The third step is to apply the personnel estimation model. For this, a linear regression model has been developed that estimates the staffing of each service that would correspond to its calculated complexity, and identifies the services that exceed or fall short of said estimate. The dependent variable considered when building the model is the total occupied personnel on 31 December 2012, and as sub-group, the inspectors (both A1 and A2). The independent variables or predictors are the total complexity of the services and the total number of entries (records) in groups of a thousand. Additionally, dummy variables are created for the qualitative variables of time and service so that they are considered when calculating the estimate. Of the models built, the one with the best statistical adjustment has been selected.

Logistic or ordinal regression

Linear regression is a regression module where the variable or fact to be explained takes either two values (the phenomenon takes place or it doesn't, i.e., yes or no) or very few values (for example a scale of 5 values that measures intensity as a lot, enough, little, or nothing). Or to put it in another way, the variable to be explained is not continuous or the function is logistic. Similar to linear regression, logistic regression allows us to adjust a point cloud to a function where an endogenous variable is partially explained through regressors.

Cost-benefit and cost-effectiveness analyses

Before assigning monetary resources to a public or private intervention, the quotient of discounted cash flows between the allocation of resources (cost) and their returns (profits) allows us to assess in absolute terms the convenience of allocating said resources or eventually of allocating them to alternative options. Occasionally, when the costs of the evaluated event are not explicit owing to the fact there is no market that reveals them, the so-called shadow prices are adopted as prices that they would have under perfectly competitive conditions.

The cost-effectiveness analysis is a variant of cost-benefit analysis that is applied when there is a lack of prices to assess the objective or set of objectives that the intervention seeks to achieve. To this end, cost would be that which allows the maximisation of the objective. When alternative interventions to achieve the same objective are compared, the selection criteria shall be to consider the intervention that helps to reach the objective at a lower cost, and at equal costs helps to maximise the objective.

Whenever faced with a problem that is resolved by cost-benefit or cost-effectiveness analysis, these techniques may constitute evaluation criteria.

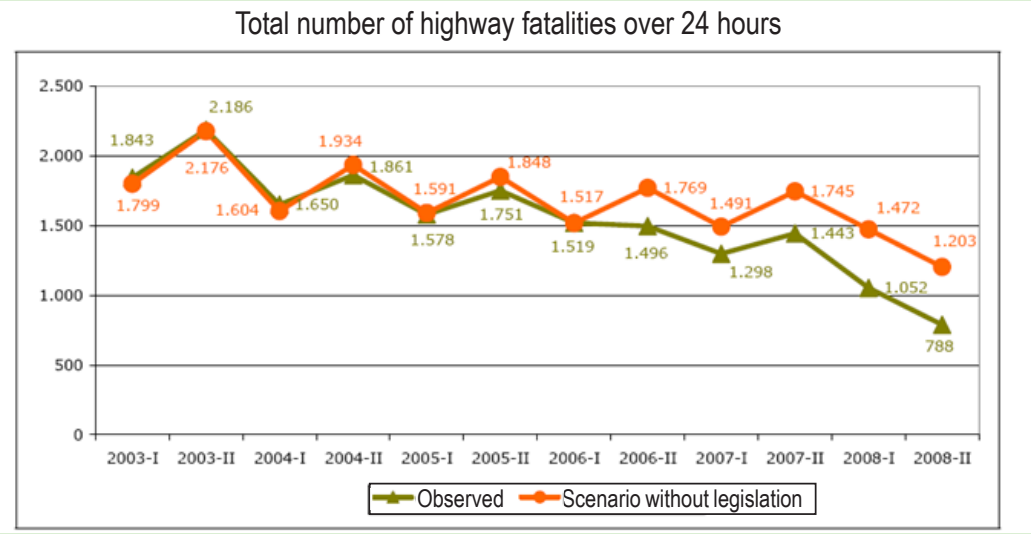
ARIMA Model

The ARIMA model (Autoregressive Integrated Moving Average) is a technique used to establish patterns of behaviour or of facts with the goal of forecasting. It does not use other variables as in regression techniques, but past data or values. Each observation is determined by earlier values in time. The ARIMA model (p.d, q) is denoted by means of three parameters - p, d, q, non-negative integers - that highlight the order of the three parts of the model: autoregression, integration and moving-average.

ARIMA models are used in evaluation to define patterns and make predictions. It is a dynamic time-series model, that is to say, future estimates are explained by the data of the past and not by independent variables.

Example: Evaluation of the Strategic Plan for Road Safety 2005-2008 (AEVAL, 2009).

This evaluation uses an ARIMA model to study the impact of certain key variables on the victims' time series, primarily the legislative changes generated by the Plan (points-based driver's license and reforming the Criminal Code, above all). The study highlights that although there was already an underlying cause that implied a descent in the number of fatalities, what is certain is that "the impact of the plan and especially, of the plan put into motion from 2006 (especially the points-based license and the Criminal Code reform) has been responsible for reducing almost all fatalities over 24 hours". The following figure displays the differences between the observed situation (green line) and that which would have occurred without the implementation of the measures according to the ARIMA model (orange line).



Multi-level analysis

Multi-level models of analysis (hierarchical linear models, linear mixed-effect and nested models, among others) are models with parameters that vary in more than one dimension. They are of use when discerning what part of an effect may be attributed to one cause and what part to another, when both are present at the same time.

For example, in research on education, they would be useful to measure what part of the students' performance is due to the teaching method or to the school of institution where they study, and what part to other variables such as the social background of the students.

Frontier or efficiency models

Another tool that helps us to analyse certain phenomena in terms of efficiency or inefficiency of the resources used with regard to the maximum potential results that may be obtained with them. These are frontier analyses of the production or cost function. Based on the definition of a Production–Possibility Frontier (PPF), these models display, firstly, the parameters that define the frontier by their functional characterisation and subsequently, the efficient options (the ones that are situated on the production frontier) and the inefficient ones. An allocation of economic resources is efficient when it is situated on the PPF.

There are three types of frontier models, of which essentially two are important:

- Nonparametric or mathematical models. This is data envelopment analysis (DEA). It uses mathematical programming to establish the set of observations that estimates the frontier and which do not require a previous functional form.

- Parametric or stochastic frontier models. It allows the estimation of the frontier functional form, costs, or benefits, the parameters, and its advantage is that it incorporates the specification error and allows us to distinguish the effects of noise or inefficiency error.

Stochastic frontier models are included in evaluation as an analytical option for applying the efficiency criteria.

Factorial analysis and principal component analysis (PCA)

When faced with a high number of variables with different degrees of correlation or linear dependency between them, both techniques may be used to reduce them to a set of factors or components that provide a synthesis of the phenomenon under study. Principal component analysis and factorial analysis both reduce the number of explicative variables, but differ in how they do it.

In the case of factorial analysis, the original variables are grouped by factors, so that they may be defined as linear combinations of the factors and explain the covariance or correlations between them.

Conversely, principal component analysis (PCA) defines new variables or linear independent components from the original variables. By means of a linear transformation, it defines a new system of coordinates for the original dataset where the highest variance is assigned to the first principal component, the second highest variance to the second component and so on, until the total variance contained in the original variables is saturated. In PCA,

components are calculated as linear combinations of the original variables, normally after centring the data in the average of each.

Both techniques may be used in evaluation for exploratory, analytical, or confirmatory purposes.

Multiple criteria analysis

Occasionally, the target of the evaluation may be assessed according to various criteria. On the basis of the weight of each criterion and according to a ratings scale, it is possible to quantitatively measure the joint application of different criteria and to sum them up in a number (the sum of the products: the weighting applied to the criterion by points attributed to the criterion), and thus compare alternatives.

Qualitative techniques

Documentary analysis

The documentation associated with the intervention is a major source of information. It refers to the documents of the intervention, programming, applicable legislation, internal orders, guidelines, budget justifications, monitoring reports, reports on results. Basically, everything that includes the institutional point of view... (Chen 1990). It is also interesting to perform a comparative analysis of the current literature and evaluations on the topic.

Interview

According to Dezin and Lincoln (2005), the interview is "a conversation, it is the art of asking questions and listening to the answers". This definition is based on a simple relationship between the researcher and the interviewee where the researcher asks questions that may range from opinion surveys or questionnaires, that is to say, highly structured instruments, to open interviews where the researcher may even be questioned or queried by the interviewee.

In qualitative research, the interview is not necessarily based on closed and structured questionnaires but on the contrary the researcher may repeat these meetings until all emerging or relevant topics have been clarified.

There are different types of interviews:

Structured interviews

In this type of interview, the questions to be asked are previously planned. A targeted and sequential list of questions is prepared. The interviewee cannot make comments or appraisals. These are closed questions; therefore the answers must be specific and exact.
Semi-structured interviews

The researcher prepares the questions beforehand on the basis of a thematic script. The questions shall be open and in contrast to structured interviews, the interviewee may express their opinions, qualify their responses, and even deviate from the initial script.
These are the most commonly used interviews in all types of evaluation, including implementation evaluations.

Figure 13. Example of semi-structured questions in the Evaluation of the measures for streamlining and improving the management of Temporary Disability. Source: (AEVAL, 2009).

Evaluation questions	Evaluation criteria
What part of the evolution of the expenditure cannot be explained by the working population, the regulatory base, or by ageing?	Suitability
Have the General State Administration (AGE in Spanish) agencies been equipped with the organisational instruments?	Coherence
Is there complementarity and coordination of the intervention between the different entities responsible for managing temporary disability due to common contingencies and has the coordination been effective?	Complementarity and effectiveness

Unstructured or open interviews

These are generally known as in-depth interviews. In this case, the objective is to "to understand the interviewees' perspectives with regard to their lives, experiences or situations, expressed in their own words" (Taylor and Bogdan, 2008). These interviews are modelled after a conversation between peers and not a formal exchange of questions and answers. They require multiple meetings with the interviewees. There are three types of in-depth interviews: life histories, learning about events and activities that cannot be observed directly, and interviewing an extensive group. These three are of great use in applied social research but not directly in evaluation, as their goals are different.

Group discussion

Group discussion is a qualitative technique which brings together a group of people to obtain information on a specific topic, conducted by an interviewer.

Group discussion is a highly valuable technique to obtain information or qualitative evidence, as it generates a series of interactions among the people who are part of the group and it aids in obtaining information that is different from what is obtained in individual interviews. When organising a group discussion, it is very important to be clear about the objective that is sought.

Based on each case, a group discussion may have different objectives:

- To share information and knowledge.
- To provide different perspectives.
- To find a common denominator.
- To come to an agreement.
- To compile qualitative information on perceptions, motivations, opinions, attitudes, etc.

There are different stages of development of a group discussion:

- Establishing objectives: The first step is to set the group objective and based on this decision, we shall define the type of group (more open or more closed) that we seek to form, whom to invite as participants (the sample), and develop tools for the group's functioning (script, schedule, activities, etc.).

- Selecting participants: In this stage we shall define the characteristics of all the participants and select the persons invited to form part of the group discussion. It is termed an "international" sample as it is not extracted on the basis of statistical criteria, nor is it a random selection, rather people are selected on the basis of their relationship with the topic under study.

- Preparing the group discussion: In this third stage, the group is planned, both with regard to the questions to be asked or the activities to be performed, and the logistic aspects.

- Group organisation: The group discussion is constituted.

- Analysing information and drawing conclusions: In this last stage, conclusions are obtained from the observations and results of the group work.

There are different types of group discussions, depending essentially on the role adopted by the group moderator and the level of conducting; from very open groups where different members of the group participate in a debate on the basis of pre-set questions, to other more focused ones that apply specific group dynamics techniques and lead the group participants towards a concrete point.

Generally, group discussions fall into two large classes:

Focus group

It is a group session, conducted by a moderator. It consists of a debate between different persons based on a list of questions that have been defined in advance and where the moderator suggests issues or asks questions and the group participants respond to them. The goal of this technique is to obtain in-depth information on a specific topic by listening to a group of persons related to the topic under analysis.

Group dynamics

In this case, we are dealing with a programmed session with a series of activities and specific group dynamics that seek concrete objectives.

The objective of this technique is highly varied, although it focuses on analysing and diagnosing, or seeking symptoms and requirements of the analysed situation. Its goal is to propose alternatives and analyse the current situation with regard to certain envisaged objectives.

This technique has its advantages and disadvantages, as it helps to pool ideas, share experiences, and build consensus. It also helps to find the common denominator between the participants. Conversely, it may lead to organisation and logistics problems and it requires prior experience. Other disadvantages are that there may arise problems, arguments, and complaints that the moderator may not be able to control.

Nominal Group Technique

Nominal Group Technique (NGT) is a creative technique for analysing problems that combines individual opinions and facilitates the decision-making process. It helps to identify the elements of a situation or problem, gives partial or total solutions to them, and establishes priorities by consulting a group of persons while respecting their anonymity.

Its development consists of five stages:

- Formulation stage. In the first stage, the questions are posed linking them to the problems, obstacles, or difficulties.
- Reflection stage. In the second stage, all participants are asked to reflect on these questions silently and individually.

- Grouping alternatives stage. In the third stage, aided by the group participants, the researcher groups all the reflections made in the first and second stages, according to the degree of similarity of each, as judged by the group.

- Debate stage. In the fourth stage, a debate is initiated on the importance of each question that has been posed. The group votes on the groups of ideas.

- Voting stage. The fifth stage corresponds to the hierarchical arrangement of the alternatives.

The process concludes with the final report drafted by the expert, who passes it on to the relevant individual or body so that they adopt the required measures and attempt to solve the problems or questions posed in the NGT, or take into account the suggestions made by the participants.

This technique has a series of advantages, among them the systematic and orderly analysis of problems, as well as making proposals for decision-making by combining individual creative responses that become qualified group opinions.

When applied to public policies, this technique allows us to identify problems and their areas of improvement. It also lets us analyse their causes and solutions. In the Evaluation of the Human Resources Quality Plan of the General Services of the Administration of the Autonomous Regions of the Balearic Islands (ACAIB) (AEVAL, 2015), this technique was applied to three groups:

- The first group consisted of nine HR managers of the General Secretariats of the council offices of the Autonomous Region of the Balearic Islands (or CAIB in Spanish).

- The second group consisted of eleven heads of all the CAIB councils with a common denominator, they had staff members and at least four years of experience in public administration.

- The last group consisted of ten ACAIB civil servants. This group was characterised by its heterogeneity.

The methodology used was common to all three groups and it unfolded in the following manner:

1. Presenting the participants.
2. Formulating the first question. In your opinion, what are the main problems that affect the management of ACAIB personnel?
3. Silent generation of ideas.
4. Collecting the ideas-responses.
5. Group discussion of the ideas-responses, interpretation, and clarification.
6. Voting.
7. Break.
8. Formulating the second question: In your opinion, how can the management of ACAIB personnel be improved?
9. Session end.

Delphi Method

It is a group technique that allows us to classify expert opinions by means of an interactive process of individual questions.

It consists of four successive rounds of questionnaires. The responses are summarised in order to draft the next consultation and an agreement is reached.

After the first questionnaire round, we come to the next stage where the experts must again respond in view of the results of the first questionnaire and justify their differences with regard to the group. In the third stage, the expert is asked to comment on the arguments that deviate from the majority, and in the last stage, a final consensus is reached. The following section displays a brief overview of the process:

- Early stage: Defining objectives, identifying interviewees, and selecting the areas of study.

- Development stage: Designing and drafting the first questionnaire. Process and obtain the average of all the results. Identifying points of divergence and homogeneity. The results of the first questionnaire are used to draft the second questionnaire and so on and so forth.

Figure 14. Delphi Questionnaires. Source: Methodological Guide on Auditing for Inspectors of the General State Administration Services (December 2009).

	QUESTIONNAIRES		
	1	2	3
CONTENT	1 or 2 open questions	Transmit, prioritise and comment in favour of / against	Transmit and revise priorities
ANALYSIS	Classify, summarise and quantify	Identify areas of agreement / disagreement and establish priorities	Establish final results

- Final stage: The results are analysed and the conclusions drafted.

The benefit of this technique lies mainly in the insistence generated by presenting the same questionnaire several times. That is to say, the results of the previous questionnaires help experts to progressively learn about the different points of view so they may continue to modify their opinion if the arguments presented appear to be more suitable than their own.

SWOT Analysis

A SWOT (Strength, Weaknesses, Opportunities, Threats) analysis is a simple and general tool for taking strategic decisions. The main goal is to help find strategic elements and use them to make changes in the organisation by consolidating strengths, minimising weaknesses, taking advantage of opportunities, and eliminating or reducing threats. This technique is based mainly on two types of analysis, internal and external.

In an internal analysis, the objective is to detect the weaknesses and strengths of the organisation: to remedy the first and to promote the second. Different aspects are studied for this purpose: production, organisation, human or personnel resources and finances.

External analysis focuses mainly on detecting threats and opportunities. For this we shall consider the environment of the organisation, interest groups, legislative, demographic, and political issues. These points are very revealing when it comes to defining strategies that seek to combat threats and take advantage of opportunities.

Once the strengths, weaknesses, opportunities and threats have been identified, the SWOT Matrix may be created, which allows us to visualise and summarise the current situation of the organisation. With the results of the SWOT analysis, a strategy must be defined.

Figure 15. SWOT Matrix. Source: Author's own.

		STRENGTHS (S)	WEAKNESSES (W)
		Of the organisation	
OPPORTUNITIES (O)	Of the environment	Take advantage of the opportunities offered by the environment, using the organisation's strengths.	Take advantage of the opportunities offered by the environment, overcoming the organisation's weaknesses.
THREATS (T)		Use the organisation's strengths to avoid the threats posed by the environment.	By reducing the weaknesses of the organisation, we avoid threats.

The advantages of this technique are mainly that it leads to an awareness of existing problems, their characteristics and how they interact with the context, the organisation or the institutional framework, as well as the risks and opportunities generated by the environment that surrounds said organisation.

LIST OF FIGURES

WORKS CITED

Zall Kusek, J., & Rist, R. (2004). Ten Steps to a Results-Based Monitoring and Evaluation System. Banco Mundial.

AEVAL. (2009). Evaluation of the measures for streamlining and improving the management of Temporary Disability. Madrid.

AEVAL. (2009). Evaluación del Plan Estratégico de Seguridad Vial 2005-2008. Madrid.

AEVAL. (2009). Evaluation of the National Plan for Transition to Digital Terrestrial Television (DTT). Madrid.

AEVAL. (2010). Fundamentos de evaluación de políticas públicas. Madrid. Obtained from http://www.aeval.es/export/sites/aeval/comun/pdf/evaluaciones/Fundamentos_de_evaluacion.pdf

AEVAL. (2011). Evaluación del Plan Español de Energías Renovables. Madrid.

AEVAL. (2011). Evaluación del Programa de Ayudas a las Actuaciones de Reindustrialización (REINDUS). Madrid.

AEVAL. (2013). Evaluación de la formación dirigida a los empleados públicos de la Administración General del Estado. Madrid.

AEVAL. (2013). Evaluación del Plan de Medidas para la Mejora de los Servicios de Sanidad Exterior. Madrid.

AEVAL. (2014). Evaluación del Plan de Ahorro y eficiencia energética 2008-2012. Madrid.

AEVAL. (2015). Evaluación del Plan de Calidad de los Recursos Humanos de los servicios generales de la Administración de la Comunidad Autónoma de las Illes Balears. Madrid.

AEVAL. (2015). Guía práctica para el diseño y la realización de evaluaciones de políticas públicas - Enfoque AEVAL. Madrid.

AEVAL. (2017). Evaluación intermedia del Plan Nacional de Desarrollo Rural (PNDR). Madrid.

Almquist, A. (2011). Attribution vs. Contribution. Centers for Disease Control and Prevention.

Apocada. (1999). Evaluación de los resultados y el impacto. Revista de Investigación Educativa, 1999, Vol. 17, n.º 2, 363-377.

Belcher, B., & Palenberg, M. (2018). Outcomes and Impacts of Development Interventions: Toward Conceptual Clarity. American Journal of Evaluation, 39(4), 478-495.

Birkland, T. (1997). AfterDisaster: Agenda-Setting, Public Policy and Focusing Events. Georgetown University Press.

Calidoni-Lundberg, F. (2006). Evaluation: definitions, methods and models. Swedish Institute For Growth Policy Studies. Working Paper.

Centers for Disease Control and Prevention. (2007). Practical Use of Program Evaluation among Sexually Transmitted Disease (STD) Programs.

CEPAL. (2005). Metodología del marco lógico para la planificación, el seguimiento y la evaluación de proyectos y programas.

Chen, H. (1990). Theory-driven evaluations. New Park, CA: Sage Publications Inc.

Chen, H. (1994). Theory-driven evaluations: Need, difficulties and options. Evaluation Practice, 15(1), 79-82.

Comisión Europea. EuropeAid - Oficina de Cooperación. (Marzo de 2001). Manual Gestión del ciclo del proyecto.

Cook, T., Scriven, M., Coryn, C., & Evergreen, S. (2010). Comteporary thinking about causation in evaluation: a dialogue with Tom Cook and Michael Scriven. American Jorunal of Evaluation, 31(1), 105-117.

Davidson, E. (2004). Ascertaining Causality in Theory-Based Evaluation. New directions for Evaluation, 2000(87), 17-26.

Denzin y Lincoln. (2005). The SAGE Handbook of Qualitative Research.

Fontaine, G. (2015). El Análisis de Políticas Públicas: Conceptos, Teorías y Métodos. Quito: Anthropos.

GAO. (2009). Program Evaluation. A Variety of Rigorous Methods Can Help Identify Effective Interventions.

García Sánchez, I. (2010). Sistema de evaluación. Obtenido de http://www.eumed.net/libros/2010b(687/

Gertler, P., Martínez, S., Premand, P., Rawlings, L., & Vermeersch, C. (2012). La Evaluación de impacto en la práctica. Banco Mundial.

Hatry, H., Van Houten, T., Plantz, M., & Taylor Greenway, M. (1996). Measurting Program Outcomes: a Practical Approach. United Way of America.

Hendricks, M. (1996). Performance Monitoring: How to Measure Effectively the Results of Our Efforts.

Instituto para la Evaluación de Políticas Públicas. (2019). Guía de evaluación de diseño de políticas públicas.

Instituto para la Evaluación de Políticas Públicas. (2019). Guía de evaluabilidad de intervenciones públicas. Madrid.

King, G., Keohane, R., & Verba, S. (1994). Scientific inference in qualitative research. Princeton: Princeton University Press.

Martinic, S. (1996). "Evaluación de Proyectos. Conceptos y Herramientas para el Aprendizaje".

Ministry of Foreign Affairs of Denmark. (2006). Evaluation Guidelines. Evaluation Department.

OCDE. (2010). Glossary of Key Terms in Evaluation and Results Based Management.
Office of Auditor General of Canada. (2009). Fall Report of the Auditor General of Canada.
Perez López , C., & Moral Arce, I. (2015). Técnicas de Evaluación de Impacto. Garceta.

Públicas, I. p. (2020). Guía de evaluación de diseño de las intervenciones públicas.

Retolaza, Í. (2018). Teoría del cambio. Una brújula para orientarte en el camino. PNUD/Hivos.

Rodríguez García, R., & Zall Kusek, J. (2004). Planning for Monitoring and Evaluating Results in Health Projects.

Rossi, P. H., Lipsey, M. W., & Freeman, H. (2003). Evaluation: A Systematic Approach. SAGE Publications, Inc.

Roth Deubel, A. (2008). Perspectivas teóricas para el análisis de las políticas públicas: ¿De la razón científica al arte retórico? Estudios Políticos, 33, 67-91.

Schalock, R. (1995). Outcome-Based Evaluation. . New York: Plenum Press.

Scriven, M. (1991). Evaluation thesaurus.

Taylor, S. J., & Bogdan, R. (2008). La entrevista en profundidad. En C. d. docencia, Métodos cuantitativos aplicados. Antología. Chihuahua.

UE. (2014). Guidance document on monitoring and evaluation. European Cohesion Fund. European Regional Development Fund.

UNPD Evaluation Office. (2002). Guidelines for Outcome Evaluators. Monitoring and Evaluation Companion Series #1.

USAID. (2010). Performance monitoring and evaluations. Selecting Performance indicators. TIPS.