



UNODC

United Nations Office on Drugs and Crime

Evaluation Handbook

Guidance for designing, conducting and using independent evaluation at UNODC



UNITED NATIONS OFFICE ON DRUGS AND CRIME
Vienna

Evaluation Handbook

Guidance for designing, conducting and using
independent evaluations at UNODC



UNITED NATIONS
Vienna, 2017

© United Nations, October 2017. All rights reserved, worldwide.

The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

This publication has not been formally edited.

Publishing production: English, Publishing and Library Section, United Nations Office at Vienna.



CHAPTER 7

Practical guide

This chapter provides practical and basic guidance for planning and carrying out quality, credible and useful evaluations. It includes common approaches and methods used in UNODC evaluations, and provides links to additional guidance and helpful resources. As the guide is focused on the fundamentals of implementing evaluations, it is most relevant to those who want to know more about what constitutes good evaluation practice, including managers of evaluations and, in particular, evaluators.

ABOUT THIS GUIDE

Evaluation is an investigative, multidisciplinary endeavour, and in order to be credible, certain principles need to be upheld, including the use of ethical standards and a rigorous methodology that adheres to UNODC requirements. This guide is not a comprehensive resource for evaluation methodology, however it does provide an overview of the fundamental requirements of good evaluation practice.

The terms of reference (ToR) provide the scope and general direction for an evaluation. However, it is up to the evaluation team to sharpen the focus of the study and determine precisely how it will be carried out. This guide aims to provide assistance along the way by addressing the main issues that the evaluation team needs to attend to in preparing and implementing a typical evaluation.

This chapter begins with a reminder of the ethical guidelines for conducting evaluations within the United Nations system.

This is followed by advice for determining the key issues to consider in an evaluation. This includes guidance on developing the preliminary hypotheses, addressing the key criteria and evaluation questions that are stated in the ToR, and ensuring that the evaluation is responsive to human rights and gender equality.

The next section addresses the overarching picture of the intervention being evaluated, specifically the theory of change and the logical framework, and the differences between them.

The chapter moves on to designing the evaluation process, including developing an evaluation matrix, determining the methodological approach, and choosing the tools for data collection and analysis. Several options for tools and brief descriptions of each are provided, although the list is by no means exhaustive. It is the responsibility of the evaluation team to develop a sound and thorough methodological approach specific to the needs of each evaluation.

The last set of guidance addresses some of the common challenges encountered in evaluation processes, and potential solutions.

The chapter concludes with a list of useful resources for evaluators. These include more in-depth information about evaluation methods, tools and technologies as well as on-line and in-person training opportunities around the globe. There are also mailing lists (listservs) to keep updated on the most recent developments in evaluation and connect with other evaluation practitioners.



ETHICAL GUIDELINES FOR EVALUATION

UNODC has clear ethical and quality standards to guide evaluation practice. All independent external evaluators undertaking UNODC evaluations are required to abide by the UNEG ethical guidelines and to sign the UNEG code of conduct. The guidelines and code are discussed in chapter two and can be accessed through the links below. Of particular relevance to evaluators is the importance of:

- Maintaining independence, objectivity and impartiality
- Respecting the rights and safety needs of stakeholders involved in the evaluation process
- Having personal and professional integrity
- Ensuring that issues of human rights and gender equality are adequately considered in the scope of the evaluation and within the evaluation process itself
- Ensuring the evaluation conforms to internationally accepted norms and standards, and the quality standards of the UNODC evaluation function
- Delivering a high-quality, timely and utility-focused product (a clear evaluation report with valid conclusions and useful recommendations)



UNEG ETHICAL GUIDELINES:
[HTTP://WWW.UNEVALUATION.ORG/
DOCUMENT/DETAIL/102](http://www.unevaluation.org/document/detail/102)



UNEG CODE OF CONDUCT:
[HTTP://WWW.UNEVALUATION.ORG/
DOCUMENT/DETAIL/100](http://www.unevaluation.org/document/detail/100)



UNODC EQA TEMPLATE: [HTTPS://WWW.UNODC.ORG/DOCUMENTS/EVALUATION/TOOLS/
EVALUATION_QUALITY_ASSESSMENTS_UNODC_EVALUATION_REPORTS.PDF](https://www.unodc.org/documents/evaluation/tools/evaluation_quality_assessments_unodc_evaluation_reports.pdf)

ETHICAL CONSIDERATIONS FOR DATA COLLECTION

It is important that steps be taken to protect the physical and psychological safety of respondents and data collectors. These include ensuring that:

- Data collection processes and tools are designed in ways that are culturally appropriate and inclusive of women and minority groups.
- Power dynamics in group settings are taken into account, and that everyone has opportunities to contribute regardless of job status, gender or other type of difference.
- Data collection visits are organized at times and places that minimize risk to respondents and that enable underrepresented groups to participate and have their voices heard.

- The anonymity and confidentiality of respondents is safeguarded (UNODC evaluators cannot identify respondents by name in reports or in any information shared with UNODC and its partners).
- All respondents are treated with respect in all interactions.

When collecting data from people who may have been directly affected by sensitive issues such as victimization, trafficking and drug use, additional steps should include:

- Having a plan in place for protecting the rights and privacy of respondents
- Ensuring the processes and tools do not create distress for respondents
- Informing interviewees about the purpose of the evaluation and asking if they agree to participate in the evaluation before they are interviewed (informed consent)
- Preparing for an emergency intervention if the participant asks for urgent or immediate help
- Ensuring interviewers are trained in collecting sensitive information and are able to provide information on how individuals in situations of risk can seek support

It is important that steps be taken to protect the physical and psychological safety of respondents and data collectors.

Further guidance on the care to be taken when interviewing people who have been victimized can for instance be found in the World Health Organization's "Ethical and Safety Recommendations for Interviewing Trafficked Women".



[HTTP://WWW.WHO.INT/MIP/2003/OTHER_DOCUMENTS/EN/ETHICAL_SAFETY-GWH.PDF](http://www.who.int/mip/2003/other_documents/en/ethical_safety-gwh.pdf)

KEY ISSUES TO CONSIDER

DEVELOPING A HYPOTHESIS

Most evaluations involve the testing of hypotheses to determine whether the predicted effects of the intervention were achieved as well as the magnitude of this effect. In many cases, the hypothesis is not explicitly stated but doing so sharpens the focus of the intervention and makes it very clear what the evaluation process is measuring. Clarifying the hypothesis is often a task for the evaluator.

A hypothesis for the evaluation formulates what the most likely explanation or theory behind an intervention is, and if and why it is effective. The hypothesis will be assessed and can be tested by various scientific methods.



A hypothesis needs to be based on a clearly stated objective that indicates that the desired change is based on certain variables. There may be a series of variables that state activities A, B and C will produce results X, Y and Z.

EXAMPLES OF HYPOTHESES FOR UNODC INTERVENTIONS

1. An intervention that aims to reach vulnerable groups who use drugs in rural areas may invest in mobile clinics to reach this particular population. The hypothesis would therefore be that the outreach programme reaches vulnerable populations and can deliver drug dependency treatment to them. An evaluation has to collect data in order to test this hypothesis and decide if there is enough evidence supporting it or if the hypothesis has to be rejected.

2. A training project that aims to improve interviewing skills of law enforcement officials when investigating human trafficking cases may have as its hypothesis that the trainees will conduct better interviews when they have participated in the training. During the evaluation process, data has to be collected accordingly in order to be able to test this hypothesis.

Additional information regarding hypothesis testing and evaluation can be found in chapter 10 “Hypothesis testing and evaluation” by C. R. M. McKenzie in the 2007 publication *Blackwell handbook of judgment and decision making* by D. J. Koehler and N. Harvey.

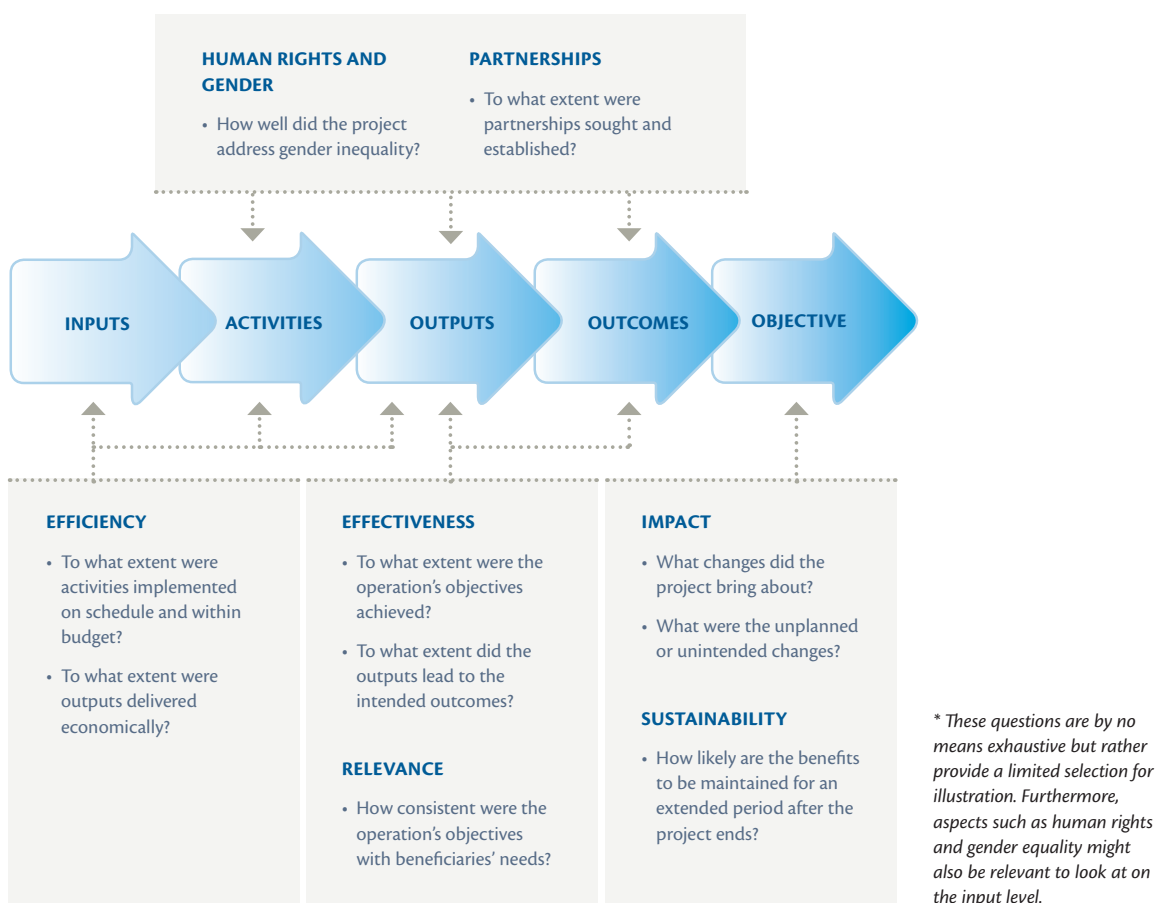
ADDRESSING THE EVALUATION CRITERIA AND KEY QUESTIONS

Next, evaluators need to take into account the OECD-DAC as well as UNEG evaluation criteria. Although not a requirement for all international development organizations, these are a standard feature of United Nations evaluations.

The criteria define the benchmarks against which to assess the evaluand. UNODC generally requires evaluators to address between six to eight criteria. Those that are mandatory include relevance, efficiency, effectiveness, sustainability, impact, and human rights and gender equality. Usually the criteria of partnerships and cooperation and, less frequently, innovation are also included. An assessment regarding the design of the intervention might be included under relevance. Descriptions of each are provided in chapter two.

The required evaluation criteria will be identified in the evaluation ToR and accompanied by key questions that the programme/project manager has determined are most important to address. Figure 7.1 shows how evaluation criteria correspond to the different levels of results in a logical framework (logframe). It also provides generic examples of questions for each criterion (the examples used are only for illustrative purposes and must not be seen as exhaustive). The criteria above the results chain, human rights and gender equality, and partnerships are cross-cutting themes that, depending on the type of intervention, can be relevant to the full range of results.

FIGURE 7.1 EXAMPLES OF EVALUATION QUESTIONS AND THE CHAIN OF RESULTS*



Source: Adapted from the IFRC (2011) *Project/programme M&E Guide*, p.14.



During the inception phase, it is critical for the evaluators to carefully review and refine the evaluation questions in the ToR and plan for how each will be answered during the evaluation.

During the inception phase, it is critical for the evaluators to carefully review and refine the evaluation questions in the ToR and plan for how each will be answered during the evaluation. It is acceptable for the evaluators to adapt, group, prioritize and sometimes even eliminate questions as long as the reasons are clearly explained and justified.

At this point, evaluators should consider the following key issues:

- Do the ToR questions adequately address all required evaluation criteria?
- Is the information that would be gained from each question of high importance to stakeholders?
- Is the number of questions manageable? Ideally, there should be no more than 2–3 questions per criterion, as having too many may result in the evaluation losing focus.
- Is it feasible for the questions to be answered accurately within the scope, timeframe and budget of the evaluation?
- Can the questions be answered by more than one source of information in order to triangulate (confirm) data?

Reaching agreement with the programme/project manager on the key questions is important before proceeding with the evaluation planning process.

We strongly discourage the use of pure Yes/No questions as evaluation questions because they are very limited in their scope and the analysis and answer that they require. Furthermore, leading questions that already implicate part of the analysis and response are also not valid evaluation questions. The same is true for multiple questions that comprise several questions in one question. Evaluation questions should be clear, well-grounded in the scope, purpose and objective of the evaluation as well as methodologically sound.³³

CONDUCTING A HUMAN RIGHTS AND GENDER-RESPONSIVE EVALUATION

Increasing emphasis and scrutiny is being placed on ensuring all United Nations evaluations address human rights and gender equality (HR and GE). All UNODC evaluations have to be HR and GE-responsive. HR and GE-responsive evaluations pay attention to the principles of non-discrimination and equality, inclusion and participation as well as fair power relations in two ways: (a) in what is examined in the evaluation, and (b) in how the evaluation itself is carried out.

³³Oxburgh, G. E. G., Myklebust, T. T., and Grant, T. T. (2010). "The question of question types in police interviews: A review of the literature from a psychological and linguistic perspective." *International Journal of Speech, Language and the Law*, 17(1), 45–66.

What does a HR and GE-responsive evaluation examine? Such an evaluation looks at the intervention's strategies, processes, practices and results. Specifically, the:

- HR and GE issues and relations that are central to the intervention
- Extent to which HR and GE were integrated (mainstreamed) into the intervention's design, implementation, and monitoring and evaluation (M&E) practices
- Progress (or the lack thereof) towards intended results regarding HR, GE and women's empowerment
- Degree to which gender relations have changed as a result of the intervention
- Extent to which the intervention has responded to and affected the rights, needs and interests of different stakeholders, including women, men, boys, girls, sexual minorities, people with disabilities, etc.

How is a HR and GE-responsive evaluation undertaken? HR and GE-responsive evaluations focus on creating space for the diversity of stakeholders involved in the intervention to engage directly in the evaluation and take some ownership over the evaluation process.³⁴ Depending on the type of intervention, stakeholder groups may include direct and indirect beneficiaries, partner organizations, as well as the line staff and senior managers of UNODC and government. At times, it may also be useful to include external stakeholders with specific expertise in human rights and/or gender such as UN Women, UNHCR, research institutions, relevant women's organizations, etc. Evaluators should aim to ensure there is diversity within each stakeholder group that is part of any evaluation process.

HR and GE-responsive evaluations are those that thoughtfully:

- Integrate HR and GE into the evaluation scope of analysis, criteria and key questions
- Use mixed, inclusive, respectful and participatory approaches, methods and tools
- Reflect HR and GE analysis in the evaluation findings, conclusions and recommendations

The importance and need for incorporating HR and GE into United Nations evaluations is discussed in chapter two. Practical guidance to mainstream a HR and GE perspective in the four phases of the UNODC evaluation process are provided in chapter four as well as the guiding document "Gender Responsive Evaluations in the work of UNODC".

HR and GE-responsive evaluations focus on creating space for the diversity of stakeholders involved in the intervention to engage directly in the evaluation and take some ownership over the evaluation process.

³⁴Miller, C. and Haylock, L. (2014). "Capturing changes in women's lives: the experiences of Oxfam Canada in applying feminist evaluation principles to monitoring and evaluation practice" (291–310). *Gender and Development* 22 (14).



The following UNEG and UNODC documents are highly recommended resources for further information and are mandatory for all evaluators who conduct UNODC evaluations. The 2011 version provides practical guidance for preparing, conducting and using HR and GE evaluations. The 2014 version provides more in-depth theoretical and practical information, tools and lessons learned.



UNEG GUIDELINES (2011):
[HTTP://WWW.UNEVAL.ORG/
DOCUMENT/DETAIL/980](http://www.uneval.org/document/detail/980)



UNEG GUIDELINES (2014):
[HTTP://WWW.UNEVALUATION.ORG/
DOCUMENT/DETAIL/1616](http://www.unevaluation.org/document/detail/1616)



UNODC GUIDANCE NOTE (2013):
[HTTP://WWW.UN.ORG/
WOMENWATCH/DIRECTORY/DOCS/
UNODC-GUIDANCENOTE-
GENDERMAINSTREAMING.PDF](http://www.un.org/womenwatch/directory/docs/unodc-guidancenote-gendermainstreaming.pdf)



LINK TO UNODC HUMAN RIGHTS
POSITION PAPER (2012):
[HTTPS://WWW.UNODC.ORG/
DOCUMENTS/JUSTICE-AND-PRISON-
REFORM/UNODC_HUMAN_RIGHTS_
POSITION_PAPER_2012.PDF](https://www.unodc.org/documents/justice-and-prison-reform/unodc_human_rights_position_paper_2012.pdf)



GUIDANCE DOCUMENT "GENDER
RESPONSIVE EVALUATION":
[HTTP://WWW.UNODC.ORG/UNODC/
EN/EVALUATION/HUMAN-RIGHTS-
AND-GENDER.HTML](http://www.unodc.org/unodc/en/evaluation/human-rights-and-gender.html)

CONDUCTING AN EVALUABILITY ASSESSMENT

An evaluability assessment examines the extent to which an intervention can be evaluated in a reliable and credible fashion. This type of assessment can be undertaken as a formal process before the actual evaluation is commissioned (e.g. already at the design stage), either by staff or by external consultants in close coordination and under the methodological guidance of the UNODC evaluation function. The purpose is to assess whether sufficient funding exists and whether the environment is sufficiently secure to carry out a good evaluation. Furthermore, it examines the existence and adequacy of a programme theory of change or logical framework (logframe), whether indicators are sufficiently SMART, and the general quality and availability of data, particularly baseline and monitoring data. If a formal evaluability assessment has not been done, these latter issues will need to be considered by the evaluator as part of the inception phase.



UNODC EVALUABILITY ASSESSMENT: [HTTPS://WWW.UNODC.ORG/UNODC/EN/
EVALUATION/NORMATIVE-TOOLS.HTML#EVAL_CRITERIA_AND_PLAN](https://www.unodc.org/unodc/en/evaluation/normative-tools.html#eval_criteria_and_plan)

THE THEORY OF CHANGE AND LOGICAL FRAMEWORK

Theories of change (also referred to as programme theories) and logical frameworks address the need to depict the cause and effect relationship between an intervention's activities and its intended results, the pathways to change.³⁵ One or both are developed during the planning stage of an intervention and provide the foundation for its evaluation. The two terms are often used interchangeably but there are notable differences between them.

A theory of change (ToC) focuses on the dynamics of change. Although there are different interpretations, ToCs are generally understood to be “a way to describe the set of assumptions that explain both the mini-steps that lead to a long term goal, and the connections between these activities and the outcomes of an intervention of programme”.³⁶

A ToC can show different levels of change, different actors and different causal pathways. It generally includes contextual factors (such as the role of the State versus civil society, the different values people hold about the topic, etc.) that help or hinder the envisioned change, and the assumptions on which it is built (conditions necessary for the change to happen but which are not under the control of the implementers). There are endless variations in how ToCs are presented. Most often they are depicted in diagram form along with a narrative description. The level of detail can vary, but simpler forms are generally more useful for sharing with a range of stakeholders.

A logical framework (or logframe) is a planning, management and evaluation tool that is based on, and reflects, the underlying ToC. All UNODC project/programme documents include a developed logframe outlining the objective, outcomes, outputs, indicators and activities of the intervention. It is usually presented as a matrix with columns that show the logical and linear relationships between an intervention's inputs, outputs, outcomes and its objective.

The focus of a logframe is typically on the results of the interventions. In order to be useful for evaluative purposes, the results need to be stated in ways that can be easily measured. For example, terms such as “enhanced or strengthened capacity” should be avoided as they do not clearly state the change that is expected. A useful resource for more information about logical frameworks is the *Results-Based Management Handbook* from the United Nations Development Group (UNDG).



UNDG RESULTS-BASED MANAGEMENT HANDBOOK: [HTTPS://UNDG.ORG/WP-CONTENT/UPLOADS/2016/10/UNDG-RBM-HANDBOOK-2012.PDF](https://undg.org/wp-content/uploads/2016/10/UNDG-RBM-HANDBOOK-2012.pdf)

³⁵http://devinfofive.info/impact_evaluation/img/downloads/Theory_of_Change_ENG.pdf

³⁶Weiss, C.H. (1995). “Nothing as Practical as Good Theory: Exploring Theory-based Evaluation for Comprehensive Community Initiatives for Children and Families.” In J. Connell, A. Kubisch, L. Schorr and C. Weiss (eds.) *New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts*. New York, Aspen Institute (65-92)



When used as a tool for monitoring and evaluation purposes, the logframe generally includes columns for the indicators that will be used to measure progress towards outcomes, available baseline data, and the associated data collection processes and sources for each indicator. Logframes are an essential component of results-based management (RBM) processes.

A section of the logframe used in the evaluation of the 2010–11 in-depth evaluation of the Global Programme Against Money-Laundering (GLOU 40) is shown in table 7.1 as an example. The complete logframe can be found in the full evaluation report, starting on page 101. A useful resource for more information about logical frameworks is the *Results-Based Management Handbook* from the United Nations Development Group (UNDG).



UNDG RESULTS-BASED MANAGEMENT HANDBOOK: [HTTPS://UNDG.ORG/WP-CONTENT/UPLOADS/2016/10/UNDG-RBM-HANDBOOK-2012.PDF](https://undg.org/wp-content/uploads/2016/10/UNDG-RBM-HANDBOOK-2012.PDF)



[HTTPS://WWW.UNODC.ORG/DOCUMENTS/EVALUATION/INDEPTH-EVALUATIONS/INDEPTH_EVALUATION_OF_THE_UNITED_NATIONS_GLOBAL_PROGRAMME_AGAINST_MONEY_LAUNDERING_PROCEEDS_OF_CRIME_AND_THE_FINANCING_OF_TERRORISM.PDF](https://www.unodc.org/documents/evaluation/indepth-evaluations/indepth_evaluation_of_the_united_nations_global_programme_against_money_laundering_proceeds_of_crime_and_the_financing_of_terrorism.pdf)

TABLE 7.1 PARTIAL LOGFRAME FOR THE IN-DEPTH EVALUATION OF GLOU 40

Specific objective 1: Legislative bodies, criminal justice officials (FIU personnel, law enforcement agencies and their personnel, anticorruption agencies officials), supervisory and regulatory authorities and the private sector aware of the negative economic and social impact of money-laundering and the financing of terrorism.	
OUTPUT	Output 3: AML/CFT policies and institutional frameworks, and possible links with existing anticorruption, good governance and antiterrorism policies reviewed and developed: Mentoring
OUTCOME	AML/CFT regulatory and supervisory frameworks developed and implemented by financial and supervisory authorities and the private sector using UNODC-provided information or services of mentors.
PERFORMANCE INDICATOR	Number of regulatory frameworks developed using UNODC information or mentoring.
SOURCE OF DATA	Reports to the CND on the number of Member States criminalizing money-laundering produced; interviews and survey results used to identify direct links between assistance provided/model laws used and amendments or laws passed.
METHOD	Content analysis; interviews with beneficiary jurisdiction; beneficiary survey results.

ToCs have increased in popularity over the years. Proponents suggest ToCs are a better depiction of the complex environment in which most development interventions take place, in particular the logical pathway of an intervention, clearly connecting activities to the envisioned effects.

A concern expressed about logframes is that they often deal only superficially with assumptions and do not make explicit the values and context of the environment in which the intervention takes place.

The logframe's focus on results is also considered by some to be problematic because the processes and relationships that are also important for success get minimal attention. For these reasons, theory-based approaches to evaluations that include a ToC are often preferred in cases where there is an interest in obtaining an in-depth understanding of the workings of an intervention.

On the other hand, well-developed logframes can include both assumptions and contextual factors, and indeed, these were fundamental components of the original logframe models. Proponents of logframes say the emphasis on results is critical, particularly in regard to evaluation purposes. They suggest that the most important issue to explore in an evaluation is the extent to which different levels of results were achieved and that there should be less emphasis on the processes involved in getting there.

Most development organizations, including UNODC, encourage the use of both logframes and ToCs. ToCs are useful because they portray the bigger picture and logframes are useful because they include the type of information needed for results-based management and evaluation processes. However, ToCs and logframes should both be:

- Based on a solid situation analysis
- Developed with the participation of key stakeholders, including beneficiaries
- Reviewed periodically and adapted as necessary in order to ensure their continued relevance.

WHAT TO DO WHEN THERE IS NO TOC OR LOGFRAME

As stated above, in order to carry out a useful evaluation, evaluators need to know what changes the intervention was supposed to make and the pathways to those changes. If the intervention does not have a ToC or logical framework that specifies the intended results in ways that are measurable, or if these are judged by the evaluators to be inadequate or outdated, then the evaluators may be asked to develop or revise at least one of these instruments. This task should be done as a consultative process that involves multiple stakeholders, in particular the project/programme teams, in order to clarify assumptions and the evaluators' understanding of the intervention.

Most development organizations, including UNODC, encourage the use of both logframes and theories of change.

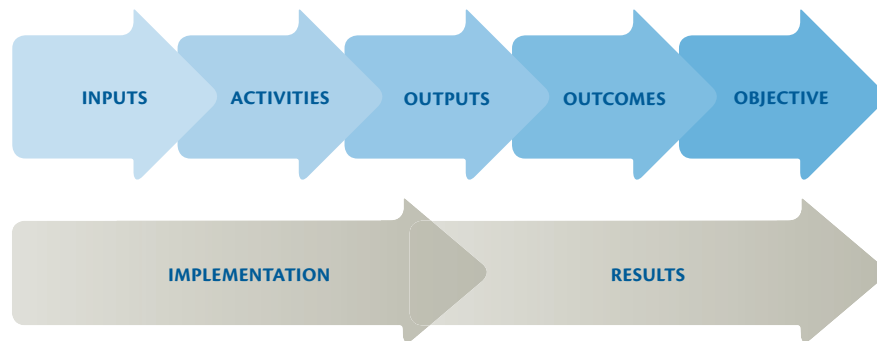


If the intervention does not have a theory of change or logical framework that specifies the intended results in ways that are measurable, or if these are judged by the evaluators to be inadequate or outdated, then the evaluators may be asked to develop or revise at least one of these instruments.

There are several ways for evaluators to proceed with reconstructing a ToC or logframe:

- Facilitate programme team meetings or stakeholder workshops for this purpose
- Review programme documents and any previous studies for information about intended and actual outcomes and impact, and important aspects of context and implementation
- Interview key stakeholders to obtain their perspectives on what the intervention is expected to achieve as well as its results to date
- At a minimum, construct a basic results chain that shows the links between what the intervention did and what the results were supposed to be. Figure 7.2 shows all of the stages of a results chain. When using this as an evaluation framework, the results chain could focus on just specifying the intervention's outputs, outcomes and objective.

FIGURE 7.2 RESULTS CHAIN



CASE EXAMPLE

A first step in conducting the mid-term evaluation of the Global Programme on Money-Laundering was to translate the objectives and outcomes in the project document into a logical framework that could be measured and therefore evaluated.

This process involved connecting the outputs and activities of the project to the outcomes that were expected to result from the activities, and organizing these in terms of the objectives that were to be achieved by 2012. The revised logframe for the evaluation was based on existing objectives and outcomes, which were converted into end-states that could be observed.

In several cases, the results were originally drafted as activities (for example, the overall objective was stated as “To assist member states in building effective legal, regulatory and law enforcement capacity . . .”) and were therefore difficult to measure. The evaluators reformulated the statements where necessary. In the case of the overall objective, it was simply restated as, “Assisted States to build effective legal, regulatory and law enforcement capacity . . .”. This is an end state on which results data can be collected.

Source: Adapted from the methodology section of the 2010–11 In-depth mid-term evaluation of GPML (GLOU40).

WHAT TO DO WHEN THERE IS NO BASELINE DATA

A baseline serves as a point of reference for measuring change over time. Baseline data are the initial data collected prior to the intervention (such as the number of arrests for drug trafficking or the laws regarding human trafficking already in place). The baseline data can later be compared to data collected during project monitoring and evaluation.

In cases where baseline data or specific data such as sex-disaggregated data for instance does not exist, it may be necessary to recreate it. The inception report should identify which of the evaluation questions require baseline data and how data collection ought to proceed in order to provide a plausible proxy for the assessment of the initial condition. To establish baselines, evaluators may conduct in-depth interviews or focus groups with key stakeholders and have them reconstruct, from memory, the original situation and the changes generated by the project. It is important for this information to be acquired from more than one source (triangulation). As already mentioned, it is also important that sex-disaggregated data is acquired and available for a proper gender-sensitive analysis.

In cases where baseline data or specific sex-disaggregated data does not exist, it may be necessary to recreate it.

DESIGNING THE EVALUATION PROCESS

GETTING ORGANIZED WITH AN EVALUATION MATRIX

A key piece in the preparation of the inception report is the development of an evaluation matrix. It is a planning as well as analysis tool used to ensure that the evaluation process addresses the key questions in a sufficiently robust manner.³⁷

Also referred to as an evaluation framework, the matrix should clearly show how data will be collected on all evaluation criteria and key questions, and how triangulation between data sources will be accomplished. The matrix is also used as the basis for designing the data collection instruments.

An example of a basic template is shown in table 7.2. The “lines of enquiry” column can be used to highlight areas that require special focus such as how to find types of data that are missing. When multiple people are developing or reviewing the matrix, it can be useful to add another column for comments.

³⁷http://web.undp.org/evaluation/documents/guidance/UNDP_Guidance_on_Outcome-Level%20_Evaluation_2011.pdf



TABLE 7.2 EVALUATION MATRIX TEMPLATE

Evaluation criteria	Evaluation questions	Indicators/subquestions to respond to each question	Collection method(s) and sources

DETERMINING THE APPROPRIATE APPROACH AND METHODS

The wide range of approaches and plethora of tools for undertaking evaluations can be both exciting and overwhelming. There is no single perfect methodology as each has comparative advantages for addressing particular concerns and needs. Evaluations have to be designed for their specific purpose, customized for the specific circumstances of the intervention, and take into account the political or decision-making context and the available budget. However, in the end, choosing the best possible approach and tools may require balancing what is ideal from a methodological perspective and what is practical and feasible.

The evaluation methodology typically addresses the broader principles that will guide how the evaluation is done. Mandatory aspects of the methodology will be specified in the ToR. For example, UNODC requires that the methodology adopts a participatory process that involves all relevant stakeholders. The need for the evaluation to be based on human rights and gender equality principles and to adhere to the UNEG norms and standards for evaluation are also standard methodological requirements for all UNODC evaluations.

The ToR may also specify types of data collection processes that are to be part of the methodology. In the end, the evaluation team is responsible for more fully developing the evaluation methodology and for clearly explaining in the evaluation report the approach, what it entails, and their rationale for selecting the specific design and methods used.

The evaluation design generally refers to the type of methodological evaluation approach that will be used.³⁸ There are many types but the most common ones for use in evaluations are:

- *Experimental*: involves the random assignment of subjects to treatment and non-treatment conditions and the pre- and post-measurement of each group
- *Quasi-experimental*: involves comparison groups and post measurements of each group to take into account the difficulties of doing a true experiment in real life
- *Non-experimental*: considers the extent of change only for those affected by the intervention and does not involve a comparison group

The type(s) of design to be used is (are) typically determined by the commissioners of the evaluation and specified in the ToR.

³⁸http://apps.who.int/iris/bitstream/10665/96311/1/9789241548687_eng.pdf

The evaluation methods are the tools, techniques or processes used to collect data. Examples of these include document review, interviews, focus group discussions, surveys, questionnaires, case studies and participant observation. Methods and how they are used are shaped by the methodology of the evaluation. Different types of data collection methods are discussed later in this chapter.

IMPACT EVALUATIONS CAN USE EXPERIMENTAL OR QUASI-EXPERIMENTAL DESIGN

Although impact can be explored and measured in various ways, the term is increasingly associated with evaluation processes that rely on evidence obtained from counterfactual analysis, with the emphasis on a valid comparison group. Impact evaluations are commissioned with the intent of objectively and rigorously measuring issues of attribution, contribution and impact.

The “gold standard” for these types of studies calls for the use of randomized control trials along with other experimental and non-experimental approaches. Well-designed and implemented impact evaluations should normally provide useful and credible information on whether the intended results have been achieved and the efficacy of the underlying programme theory.

However, such studies generally require the implementation of strong monitoring systems designed to gather impact information. They are also usually costly to undertake due to the need for large amounts of data as well as the considerable amount of time and resources to collect data and ensure its quality.

The following World Bank publication by P. J. Gertler, S. Martinez, P. Premand, L. B. Rawlings and C. M. J. Vermeersch from 2011 is a useful resource for impact evaluations: https://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf

Evaluations are considered more credible, and therefore are more useful, when the methodology is sound and rigorous. A sound approach includes the use of valid and reliable methods of data collection and analysis, and the use of triangulation for developing the findings.

VALIDITY AND RELIABILITY OF DATA

Evaluations are considered more credible, and therefore are more useful, when the methodology is sound and rigorous. A sound approach includes the use of valid and reliable methods of data collection and analysis, and the use of triangulation for developing the findings. In addition, the methodology has to adhere to validity as well as reliability.³⁹

- *Validity* refers to the accuracy and relevance of the data, i.e. whether or not the data collection tools are measuring what they are intended to measure.
- *Reliability* refers to having consistency in results using the same method, i.e. whether similar findings would come from using the same tool multiple times.

³⁹<https://www.oecd.org/development/evaluation/qualitystandards.pdf>



It is a good and necessary practice to test all data collection tools to ensure high validity and reliability.

There are multiple methods for ensuring that data collection tools exhibit high validity and reliability. For example, to ensure validity, information obtained from using a tool multiple times should be reviewed to ensure it corresponds to the evaluation question. To ensure reliability, the tool could be tested more than once on the same individual; the tool could be administered by several people; or the tool could contain multiple questions that are aimed at answering the same question. It is a good and necessary practice to test all data collection tools to ensure high validity and reliability.

Another important concept for ensuring the credibility of data is triangulation. Evaluation findings are strengthened when multiple pieces of evidence point in the same direction. Triangulation is accomplished when different data or types of information are used that come to broadly the same conclusion.

- *Triangulation* refers to using multiple approaches, methods and sources for data collection and analysis to verify and substantiate information.⁴⁰ It enables evaluators to overcome the bias that comes from single informants, methods, observations or points of view. The purpose of triangulation is to increase the reliability of evaluation findings.

TRIANGULATION ILLUSTRATED

An example of triangulation in a UNODC context is found in the 2016 independent project evaluation of the AIRCOP programme.* Thorough triangulation was achieved by combining several methods of data collection and by including different stakeholder groups as informants in the evaluation.

The evaluation team conducted thorough desk research, situational analyses of the regions where the programme implemented activities, 116 interviews with different stakeholders and core learning partners, observations in the field, as well as an online survey. This sound evaluation methodology enabled the evaluation team to triangulate their findings by method, source as well as by evaluator.

* https://www.unodc.org/documents/evaluation/Independent_Project_Evaluations/2016/XAWU72_AIRCOP_mid-term_evaluation_report_September_2016.pdf

Rigorous methodologies are generally characterized as those that use sound approaches following strict standards. Although evaluations carried out using experimental designs are often held up as being the most rigorous, rigor is also essential to qualitative approaches and is also achieved by using a mixed methods approach.

⁴⁰UNEG Norms and Standards 2016.

QUANTITATIVE, QUALITATIVE AND MIXED METHODS⁴¹

Evaluation studies can use a quantitative, qualitative or mixed-methods approach, as the evaluators try to find a workable balance between procedures that ensure the validity of findings, and those that make findings timely, meaningful and useful to those who commission evaluations.

The terms qualitative (QUAL) and quantitative (QUANT) can be used to describe an evaluation design, method and type of data. Both approaches have advantages and disadvantages; they each provide important evidence to an evaluation process but neither is without bias.

Quantitative approaches are experimental or quasi-experimental and answer the “what” questions. They measure and explain what is being studied with numbers. At a basic level these measurements give counts, ratios and percentages.

QUANT methods tend to use structured approaches that provide precise data that can also be statistically analysed and replicated for comparison. Statistical analysis provides for a more complex understanding of the data, such as a comparison of means, comparison of differences, and making predictions.

Although QUANT data is considered to be objective, there are subjective elements to designing QUANT processes, such as determining what variables to measure, what questions to ask, and the ways in which results are analysed and reported.

Qualitative approaches are non-experimental and answer the “why” and “how” questions. They analyse and explain what is being studied with words (notes taken of respondents’ perceptions, documented observations). They focus on the constructed nature of social programmes and consider context, perspectives and values as part of determining the results of an intervention.

QUAL methods use semi-structured techniques to gather data that potentially provide in-depth understanding of attitudes, beliefs, motives and behaviours. They tend to be more participatory and reflective in practice than QUANT methods.

QUAL methods are primarily subjective but should be designed in ways that provide objective and quantifiable data. The rigor of QUAL approaches is achieved through use of structured and systematic data analysis processes (such as content analysis), and incorporating practices such as inter-rater reliability and self-reflection.

A mixed-methods evaluation approach is recommended for all UNODC evaluations.

⁴¹Cresswell, J. W. (2014). *A Concise Introduction to Mixed Methods Research*. London: Sage.



As it is usually not feasible to obtain information from all stakeholders or locations affected by an intervention, sampling is used to enable the evaluator to generalize findings to a broader population.

A *mixed-methods* evaluation approach is recommended for all UNODC evaluations. This allows evaluators to utilize the advantages of both, measuring what happened with QUANT data and examining and why it happened with QUAL data.

THE IMPORTANCE OF SAMPLING

Most evaluations will need to use some type of sampling process in order to be considered credible. As it is usually not feasible to obtain information from all stakeholders or locations affected by an intervention, sampling is used to enable the evaluator to generalize findings from a representative selection to the broader population.

Sampling can be employed for a number of purposes: to select respondents, destinations for field visits, projects to observe and documents to review. There are two broad categories of sampling approaches: probability and non-probability (or purposive sampling).⁴² Examples of each are provided in table 7.3.

With probability sampling, each case or member of the target population has a known probability of being selected. If there is a need for the sample to be statistically representative of the total target group (universe), as would be the case in impact evaluations, the number of units required in the sample must be calculated. Sampling errors (the degree to which the sample might differ from the total population) can then also be derived.

With non-probability sampling, a purposeful process is used to select relevant and information-rich cases. It is generally up to the evaluator to determine the size of a purposeful sample. It needs to be large enough to be credible given the purpose of the evaluation.

Any probability sampling, if it is well done and therefore if it is statistically representative, will have a balanced gender representation. In the case of non-probabilistic sampling the sampling strategy has to ensure a gender balanced representation and, if it is not possible, explain why as well as outline mitigating factors.

⁴²http://www.unaids.org/sites/default/files/sub_landing/files/10_4-Intro-to-triangulation-MEF.pdf

TABLE 7.3 APPROACHES TO SAMPLING

PROBABILITY SAMPLING	
Random sampling	Selection is completely by chance, with all units having the same likelihood of being chosen. This approach avoids selection bias. If the sample size is relatively small, a simple process to select a random sample is to write all possibilities on individual pieces of paper. Once the number of units to be chosen is known, that number of papers can be chosen from a pile without looking at them. Random sampling is a feature of evaluation processes using an experimental design.
Stratified random sampling	Used when there are different groups that you know need to be represented in your sample (based on location, gender, stakeholder group, budget size or other attribute). Applies random selection of all units within each group. This is a feature of quasi-experimental design, and also addresses concerns of selection bias.
NON-PROBABILITY SAMPLING	
Maximum variation sampling	This aims to capture cases that represent the most diverse characteristics of the universe, i.e. the sample includes the cases where the intervention was the most successful and where it was least successful.
Opportunistic sampling	When new opportunities arise in the course of data collection, these are added to the sample. Such situations may occur during field visits when evaluators come across individuals or situations they feel are relevant and useful for inclusion in the study.
Snowball sampling	A strategy that aims to discover new informants who might be particularly useful to the study by asking the people interviewed for others who might know about a particular topic. It is helpful for finding key persons who are otherwise not known to the evaluator.
Purposive sampling	This sampling technique relies on the researcher's/evaluator's judgement as to who to include as a study participant or interviewee from a certain population. This might be appropriate when only a limited number of people can serve as primary data sources. It is cost and time effective, but can however introduce a high level of bias.
Convenience sampling	This means that informants are included due to their accessibility and proximity. Representativeness is not carefully considered and this might introduce a sampling bias and limits to generalization. This strategy is easy, fast and inexpensive.

METHODS AND TOOLS FOR COLLECTING DATA

There are many options for gathering data and interesting tools and technologies are constantly being developed. The use of innovative methods is encouraged but choices need to take into account the challenges and realities of the context in which they will be used. Tools and technologies also have to be adapted for the purpose and scope of the evaluation.



Evaluators are required to use multiple tools for triangulation purposes but are also encouraged to select a varied mix of methods and tools in order to get a richer set of data.

Tools are often assumed to be either QUAL or QUANT, but most often the line is blurred. Tools that are categorized as QUAL can often have QUANT elements, such as interviews that include ranking questions. QUANT tools typically include what are considered to be QUAL elements, for instance in form of open-ended questions in questionnaires. In addition, most methods require some level of subjective judgment that is based on the evaluators' interpretation of the context and therefore introduce an element of bias. This is the case, for example, in the choice of questions that are selected to be included in surveys.

Evaluators are required to use multiple tools for triangulation purposes but are also encouraged to select a varied mix to get a richer set of data. It is all too common to restrict data collection to document review and key informant interviews. But there is a far greater range of QUAL tools, in particular, which can be used to provide vital context information and that often produce more robust findings.

Table 7.4 includes 17 different methods that are commonly used and relevant for UNODC evaluations. The methods are listed in alphabetical order and not in order of relevance. The resources cited at the end of the chapter provide more in-depth information about these methods and many more method options to choose from.

Please note that the methods refer to both primary and secondary types of data. If the data is collected for the first time directly by the evaluation team it is considered primary data. Secondary data is data collected by using already available sources (usually in document form).⁴³

TABLE 7.4 METHODS FOR DATA COLLECTION*

Case study	A detailed description of a limited number of observations (i.e., of a community, work team, project, time period, etc.). Case studies are useful for evaluating complex situations and exploring qualitative impact. This approach generally involves multiple cases for comparison purposes in order to see commonalities. The evaluator looks at patterns of data in order to identify main issues that emerge, triangulates key observations, sees if there are alternative interpretations to pursue, and is then able to make generalizations about the cases. Comparing experiences of high-, medium- and low-performing cases can help to identify what factors made a particular intervention successful.
Checklist, rating scale, scorecard	Tools that state specific items or criteria to help an evaluator see the extent to which outcomes and standards are being met. They allow for data to be collected in a systematic way, and can be useful for enabling data to be expressed in a quantitative way.

*Adapted from the IFRC (2011) *Project/programme M&E Guide*, p.101-102

⁴³https://www.unicef-irc.org/publications/pdf/brief_10_data_collection_analysis_eng.pdf

Document review	Can include a wide range of material that has relevance to the intervention being studied. This can include internal documents (programme/project reports and records, prior reviews and evaluations, training materials, policies, administrative databases, videos, photos) and externally produced documents (reports and surveys from national governments, research institutions and other development organizations). Such documents are considered secondary sources of data. They can provide a cost-effective way of collecting baseline data, context, and a historical perspective of the intervention.
Expert panels	A variety of experts engaged when highly specialized input and opinion is required. The experts typically represent different fields. They are brought together during an evaluation in real-time, via on-line discussion forums, or e-mail exchanges to debate and discuss various courses of action and make recommendations. The experts should be identified by, or in consultation with, the programme/project manager and CLP.
Focus group discussion (FDG)	A discussion undertaken with a small group of participants (preferably fewer than 12) to obtain perspectives and beliefs relevant to the issue being examined. A facilitator uses a prepared list of question areas to guide the discussion but not to control it. It is important that facilitators use techniques that enable all participants to contribute to the conversations. In contrast to group interviews, the aim of FDGs is for participants to discuss and debate issues with each other with the facilitator taking the role of guide, observer and recorder. It is highly recommended to have another person taking notes.
Interview	A standard method in all evaluations that can be conducted on an individual or group basis. A structured (closed-ended) interview follows a structured set of pre-prepared questions that generally only allow for a limited range of answers (such as yes/no or expressed by a number/rating on a scale). Responses can then be coded for statistical analysis. A semi-structured (open-ended) interview allows for more in-depth responses to questions. The interviewer still has pre-prepared questions but these are for guiding the discussion, and the interviewer generally has the discretion to probe issues of particular interest more deeply.
Key informant interviews	Done with people selected because they have specific or specialized information about a particular topic. The interviews typically follow an open-ended format.
Most significant change (MSC)	A participatory technique whereby participants are asked to describe the most important change that has happened from their perspective as a result of the intervention. MSC is often used when no, or only limited, baseline data or indicators exist. The stories that emerge can provide a rich picture of the impact and can provide the basis for further discussion of the intervention's value and programme theory.

(cont.)



Participatory methods	A wide variety of techniques that actively engage participants and generate open discussions that often bring out a wider range of ideas and perspectives than more conventional forums. They are also useful for starting conversations about challenging topics and building relationships between the evaluators and respondents, as well as between respondents. Examples include ranking exercises and having participants draw diagrams, maps, timelines and other visual displays to examine the study topics. The techniques can be particularly effective in situations with mixed-language or low-literacy groups, where not all participants would otherwise feel comfortable contributing ideas, where the desired information is not easily expressed in words or numbers, or where a group that may have grown tired from a more structured question and answer session needs to be energized. Participatory methods require sufficient time for planning as well as for the actual implementation.
Participatory rapid (or rural) appraisal	A process used to engage communities in an evaluation. Community views on a particular issue are gained in an intensive manner over the course of up to two to three weeks through a series of methods such as focus groups, community mapping and interviews.
Observation	Generally involves spending considerable time observing events, processes or people as they go about their typical activities, and recording these observations. In the case of participant observation, the evaluator interacts with people as a participant in their community or group. This technique is useful for gathering insights that may be missed in more structured forms of data collection.
Questionnaire	A data-collection instrument that contains a set of systematically organized questions. It typically includes a set of instructions about how to ask the questions so that the data collectors/interviewers implement the questionnaire in a standardized way. The questions can be aimed at eliciting quantitative data but also qualitative data in the form of free answer formats.
Statistical data review	A review of existing data from sources such as research studies, government-generated census information, etc.
Story	An account of an individual's experiences and how that person has been affected by the intervention being reviewed. Typically, some background information is provided in order to provide context for the account. The stories are usually framed as success stories, which are based on interviews with participants and focus on the positive impacts the intervention has had on participants' lives.
Survey	<p>A set of questions designed to systematically collect information from a defined population, usually by means of interviews or questionnaires administered to a sample of people representative of the target population. An enumerated survey is one which is administered by a trained data collector. A self-administered survey is completed by the respondent.</p> <p>On-line surveys are a common and cost-effective method of collecting data in UNODC programme-level evaluations, particularly from staff as well as training participants. However, it can be difficult to get a sufficient response rate in order for the results to be useful. Techniques for increasing the response rate include keeping the survey short, using clear and concisely worded questions, ensuring questions are relevant to and in the language of the target group(s), sending multiple reminder notices to targeted respondents, and having managers encourage their teams to complete the survey.</p>

RANKING: A SIMPLE AND ADAPTABLE PARTICIPATORY METHOD*

Participatory techniques might be challenging for some evaluators. However, they are widely accepted and particularly useful for unpacking complexity. For those wanting to expand their toolkit, ranking is one example of a participatory method that can be easily applied to almost any evaluation process.

There are many ways of ranking or ordering information. For example, performance ranking can be done with staff to understand how they rank specific elements of a project: what worked, what is not working so well, and what improvements could take place. Respondents can also be asked to brainstorm a list of challenges they face and then vote on which are the most important to be addressed for subsequent phases of the project (i.e. making the work environment more supportive for women).

* http://devinfolive.info/impact_evaluation/img/downloads/Participatory_Approaches_ENG.pdf

HARNESSING THE POWER OF BIG DATA FOR EVALUATION*

What is big data?

In essence the concept refers to the collection and analysis of large volumes of data through innovative means and the integration of a diversity of sources of data. It aims to discover patterns in data and predict outcomes of interventions using passive acquisition methods to collect data and algorithmic and machine learning techniques to analyse it.

Why is it useful?

The world is flooded with data. It is estimated that 90 per cent of the data in the world today has been created in the last two years and the majority of this data is never analysed. The possibility to transform a wealth of disorganized data into actionable evidence is one of the potential benefits that big data brings to the evaluation field.

How can UNODC benefit from using it in evaluation?

UNODC frequently operates in complex settings with data paucity where traditional methods of data collection are neither possible nor advisable. The passing of a law against corruption in a specific country as a result of UNODC initiatives by providing legal advice and technical assistance to legislative operators triggers a myriad of cause and affects relations (e.g. awareness, budget allocations, changes in the criminal code, etc.).

Capturing different strands of data from different social media platforms, national statistics and third-party monitoring systems could show changes that UNODC intervention alone would not be able to generate. In this specific case, variables such as the perception of the population on corruption, reduction or increase of offences on corruption, etc. before and after the passing of a law fighting corruption, could help to estimate UNODC contributions to transformative change.

* http://unglobalpulse.org/sites/default/files/IntegratingBigData_intoMEDP_web_UNGP.pdf



Carefully recorded notes make it easier to triangulate data and data sources.

Keeping track of data sources

It is important to keep an ongoing record of all sources of data. UNODC policy does not permit the names of respondents to be reported, however other types of data on sources need to be included in evaluation reports and records. Data collection instruments should include space to record stakeholder group, organization represented, gender of respondent and other information such as age, as identified in the evaluation matrix or data collection plan. This will enable the evaluator to provide disaggregated statistics on respondents in the methodology section of the evaluation report. The documents reviewed should also be tracked. Listing these according to categories as they are used will ease the process of pulling together the final report.

Careful recording of data

Taking good notes throughout the data collection phase reduces the risk of losing potentially valuable data and eases the report writing process. Detailed records should be kept of what is said during interviews, discussions and debriefing meetings with other evaluation team members.⁴⁴ As such notes are often taken hastily, it is advisable to review, improve and transcribe (if recorded) them the same day.

The memories of these sessions will always be selective and there is a risk of losing crucial information. Translations should also be done as soon as possible in case any information needs to be clarified with the translator. It is also helpful to keep a field diary where insights gained and questions that arise can be recorded on a daily basis.

Although the above may be stating the obvious, field trips to collect data are typically very tightly scheduled and there are many logistical issues to attend to, which take time. It is good practice to block out part of each day for reviewing notes.

In addition to the need for accurate records when it comes to writing the report, it is important to keep in mind that these notes constitute raw data that may need to be referred to later. Carefully recorded notes make it easier to triangulate data and data sources.

CODING QUALITATIVE DATA TO PREPARE FOR ANALYSIS

Qualitative data collected from document review and field missions needs to be organized in a way that allows for thorough data analysis. Coding is the process of systematically organizing text-based information (from field notes, interviews, documents) into usable empirical data, in other words, breaking down large amounts of text into manageable categories.⁴⁵

⁴⁴Sutton, J. and Austin, Z. (2015). "Qualitative research: Data collection, analysis, and management," pp. 226-231. *The Canadian Journal of Hospital Pharmacy* 68(3).

⁴⁵Ibid.

A code is a label to tag a concept or value found in the text, usually related to the outcome indicators (such as key words and themes). The coding process can be done in steps that progressively develop unsorted data into more refined categories, themes and concepts. Numerical codes can be assigned to produce quantitative statistics.

EXAMPLE: CODING OF INTERVIEW DATA

It is good practice to have two evaluators conducting each interview. This allows for one person to be the main note taker and the other to be the main interviewer. The second evaluator can then check and complement the notes that were taken.

The notes should then be coded and categorized in accordance with the evaluation questions and regarding their relevance for answering different aspects of the evaluation criteria. The coding process should also be conducted by more than one person in order to minimize bias, increase objectivity and inter-rater reliability.

The process of data analysis should be led by exchange and discussion among the evaluation team as well as a process of consensus finding. The systematic and accurate data coding process will enable the evaluation team to revisit the data throughout the evaluation process in order to reassess certain questions or provide evidence of careful data triangulation.

The inception report should lay out a clear plan for how different types of data collected will be organized, triangulated, analysed and presented in the findings section of the evaluation report.

MAKING SENSE OF THE DATA THROUGH ITS ANALYSIS

Data analysis involves systematically identifying trends, clusters or other relationships between different types of data. This aggregation and synthesizing of evidence is an essential part of the evaluation process.⁴⁶ The inception report should lay a clear plan out for how different types of data collected will be organized, triangulated, analysed and presented in the findings section of the evaluation report.

The data analysis phase is also time for internal quality assurance, keeping in mind that the quality of evidence is the backbone of a credible evaluation.⁴⁷ The evaluators should assess the strength and validity of the evidence that has been collected, and see if there are gaps or shortcomings that need to be addressed.

⁴⁶https://www.unicef-irc.org/publications/pdf/brief_10_data_collection_analysis_eng.pdf

⁴⁷ Donaldson, S. I. (2014). "Examining the backbone of contemporary evaluation practice: Credible and actionable evidence." In S. I. Donaldson, C. A. Christie, and M. M. Mark (Eds.). *Credible and actionable evidence: The foundation of rigorous and influential evaluations*. Newbury Park: Sage.



The quality of evidence is the backbone of a credible evaluation.

All data collection and analysis processes have some type of limitation and this needs to be reported. Data analysis is always based on interpretation of material drawn from sources that may have unreliable elements. For example, samples may not be representative, and there may be information gaps in key documents or data sets. It is critical to take note of these factors and clearly explain the limitations of the study in the Methodology section of the evaluation report. However, it is also similarly important to outline the mitigating measures that have been developed and implemented in order to address any limitations.

Table 7.5 provides a brief overview of different methods that can be used for analysing and interpreting different types of data. They ensure varying levels of robustness. More information about these and other methods can be found in the resources section at the end of this chapter.

TABLE 7.5 METHODS OF DATA ANALYSIS*

Benchmarking	Compares the performance of an organization's interventions with similar work conducted by other development organizations working in similar environments. The challenge of benchmarking is that rarely do organizations have the same processes and methods for establishing indicators and collecting results data, which makes direct comparisons challenging.
Content analysis	A common approach to analysing qualitative data. The data, once coded, is reviewed and analysed for trends, patterns, similarities, etc. Coding and content analysis can be done by using software programmes such as NVivo, Atlas, RQDA, etc., word processing applications, databases such as Excel and Access, or by hand, i.e., using different coloured markers to highlight different themes in text documents. There are both deductive and inductive approaches to content analysis. Deductive approaches are those that look to confirm or reject a hypothesis. Inductive approaches are more exploratory and look at the range of individual data in order to make broader generalizations.
Cost-benefit analysis (CBA)	An approach for assessing efficiency by calculating and comparing the positive and negative consequences of an intervention in monetary terms. Although data on the financial costs can be readily obtainable, assigning value to benefits is more complex. CBA uses methods to assess people's willingness to pay for the benefits (often intangible) they will receive as the result of an intervention. As CBA has limitations in terms of how people assign values to different items such as social benefits, it is often best used as part of a multi-criteria analysis.
Cost-effectiveness analysis	An approach for assessing whether results are being achieved at a reasonable cost. Also used to assess efficiency, it typically considers the cost per unit of a service given or the cost per beneficiary. The usefulness of this approach is increased when unit costs can be compared with other similar interventions.

Multi-criteria analysis	A set of methods that address cost and benefits of an intervention that take into account monetary values as well as non-monetary values relevant for a successful intervention such as time savings, project sustainability, and social and environmental impacts.
Qualitative data base	A system for organizing data by categories, particularly useful for arranging evidence from document review by outcome-level results. This can be done using a basic Excel spreadsheet.
Statistical analysis	A way of summarizing and analysing quantitative data, usually obtained from surveys. Descriptive statistics are used to understand characteristics of the sample studied (i.e. average age, percentage exhibiting a certain behaviour, income range). Inferential statistics are used for testing hypotheses and drawing conclusions about a larger population set, based on the sample, by using basic processes such as T-Tests, confidence intervals and regression (i.e. percentage of border control officers likely to implement new behaviour from training if they take a supplementary e-learning module). Descriptive statistics are automatically generated by many on-line survey tools. Excel or more robust statistical packages such as SPSS and SAS can be used for both inferential and descriptive analysis.
SWOT analysis	A basic method of assigning qualitative evidence into the four broad categories of strengths and weaknesses (internal factors which project implementers have some control over) and opportunities and threats (external factors that can significantly affect project success). It can be used as a participatory tool for gathering data, whereby respondents are asked for their perspectives on an intervention according to each category, and as a tool for evaluators to draw conclusions from evaluation findings.

*Josselib, J.-M. and Maux, B. (2017). *Statistical tools for program evaluation*. Springer; https://www.portugal2020.pt/Portal2020/Media/Default/Docs/AVALIACAO/02-Evaled_Sourcebook_Set2013.pdf

COMMON CHALLENGES

COMMON RISKS AND NECESSARY ACTIONS

As with any other evaluation process or study/research approach, there are certain challenges implicit in UNODC evaluations that need to be outlined in order to increase awareness from the beginning as well as strengthen the provision of mitigating measures. The challenges can be broadly categorized in the following four areas: (a) evaluation process; (b) expertise of the evaluation team; (c) methodology and (d) UNODC topics. Each area comprises different aspects that pose a risk to or are challenging for the UNODC evaluation process if due consideration and careful planning is not paid. The four areas are outlined in more detail below.



Evaluation process

Aspects that could be a challenge for the evaluation in this category are the resources allocated for the evaluation process as well as the time designated for certain steps of the evaluation process. In addition, the inclusive outreach during the overall process could be another challenge.

Resources. As outlined repeatedly throughout the handbook as well as emphasized in the UNODC evaluation policy, it is essential for a successful and quality evaluation to reserve sufficient budget when planning an evaluation.

Time. In particular, two aspects are essential regarding the issue of time. On the one hand, the planning and initiation of the overall evaluation process has to be started way in advance in order to allow for adequate time for all steps of the evaluation process and successful coordination with the evaluation function. On the other hand, enough time has to be allocated for each of the steps in the evaluation process, especially the inception phase, data collection and writing of the report. Furthermore, the review phases with several rounds between the evaluation team and the UNODC evaluation function also have to be considered when drafting the time frame for the evaluation.

Inclusiveness. The UNODC evaluation function requires inclusive, participatory evaluation approaches. Therefore, strong CLP engagement as well as outreach to as many stakeholders as possible are encouraged. In addition, different perspectives and underrepresented groups should also be included and considered in the evaluation.

Expertise of the evaluation team

Another challenge for UNODC evaluations is the recruitment of highly experienced and qualified external evaluators who conduct the evaluations. The high expectations as well as requirements placed on the consultants hired for the evaluation team, constitute yet another challenge for the overall evaluation process.

Evaluation expertise. The consultants hired to be part of the evaluation team should be highly experienced in developing sound evaluation approaches and conducting evaluations. Evaluators working with UNODC are required to have several years of substantive evaluation experience and knowledge.

Substantive expertise. In addition to the evaluation expertise, substantive expertise in the topic to be evaluated is expected from at least one of the team members of the evaluation team.

Expertise in human rights and gender mainstreaming. Expertise or sound experience in human rights based and gender equality-responsive evaluation approach is expected from at least one of the evaluation team members.

Methodology

Data collection tools. Data collection tools have to be carefully developed based on the evaluation questions and in particular strongly anchored in the evaluation matrix in order to fulfil the requirements of the evaluation ToR and collect data for thorough and sound data analysis, which is the basis for the evaluation report.

Triangulation. All data has to be rigorously triangulated in order to ensure thorough data analysis. Thus, the findings, conclusions and recommendations are based on triangulated data. Triangulation should be done by theory, source, method as well as evaluator.

Confidentiality. Participants who take part in the evaluation as informants, survey respondents or interviewees have to receive assurance that their information will be treated confidentially. In addition, the report will not provide identifiable information but findings will be reported anonymously, relating to group findings only.

Survey response rates. Many UNODC evaluation approaches include a survey component in addition to face-to-face or phone interviews. This practice is on the one hand beneficial in reaching a wider stakeholder group and asks for information in a more systematic and often quantitative way. However, on the other hand the response rate to surveys usually remains a challenge. Continuous and close follow-up of potential respondents as well as a sophisticated design and flow of the survey are essential in attracting relevant responses.

UNODC topics

Sensitive topics. The thematic areas UNODC engages in are of course often sensitive, for instance terrorism prevention, criminal justice reform or human trafficking. These issues often involve aspects of national security as well as involvement of the criminal justice system. For the evaluation, this means that sometimes data is protected and not easily available; counterparts are reluctant to share their experience and projects sites are difficult to access. These factors need to be considered when planning the evaluation and especially during the field missions of the evaluation team.

Gender neutrality. Some thematic areas of UNODC might be considered neutral to gender. However, there is no such thing as a gender-neutral theme, rather we have gender-blind areas of practice. Thus, gender-blind interventions risk perpetuating and reinforcing existing patterns of discrimination and exclusion. A gender-responsive evaluation approach as followed and emphasized by UNODC will contribute to assessing all thematic projects and programmes in accordance with the principles of inclusiveness, participation, gender equality and the empowerment of women.



Marginalized groups or groups in a vulnerable situation. Another factor inherent in the field of UNODC mandates is that it often involves groups in a vulnerable situation such as for instance victims of crimes, vulnerable prison populations or marginalized groups with less access to the legal system. The evaluation has to be aware of these rights holders, involve them in the evaluation process and also evaluate the programme/project with regard to the effects that it has on them. The human rights-based and gender equality-responsive evaluation approach should ensure that this aspect is seriously considered and the evaluation adapted accordingly.

Hidden populations. Some of the stakeholders of a programme/project or an intervention might be hidden populations that are difficult to reach. The evaluation team has to be aware and considerate of these hidden populations and has to develop creative, inclusive approaches in order to recognize their perspectives and include the information obtained from them in the data analysis.

Overrepresentation of men in law enforcement and government positions. Globally, men outnumber women as judges and magistrates. Furthermore, women make up fewer than 35 per cent of police personnel in countries with available data.⁴⁸ Thus, the gender equality-responsive evaluation approach required for all UNODC evaluations aims at capturing gender perspectives and the voices of women and other marginalized groups, even though originally men might be overrepresented as evaluation stakeholders in the initial evaluation approach. The evaluation teams are strongly encouraged to reach out to female stakeholders, minorities and marginalized groups as well as other United Nations organizations and CSOs in order to arrive at a representative analysis of the relevant context for a human rights and gender perspective on the topic under evaluation.

The challenges outlined above clearly illustrate the complexity of UNODC topics that are being evaluated. This complexity is another challenge implicit in the thematic focus of the work of UNODC. The aspect of complexity will be outlined and discussed in more detail in the section below.

ADDRESSING COMPLEXITY⁴⁹

Why is complexity important?

Whether in professional settings or in a personal context, we are all surrounded by complexity. As managers, decision makers or evaluators we all contribute and suffer the effects of complexity. The picture of a butterfly flapping its wings in the Brazilian rain forest being the original cause of a chain of events triggering a hurricane in the northern hemisphere is a metaphor that summarizes the importance of small actions interacting with diverse elements part of a larger system (the butterfly as part of the biosphere as part of the weather system).

⁴⁸ <https://unstats.un.org/unsd/gender/worldswomen.html>

⁴⁹ <http://www.fsg.org/publications/evaluating-complexity>. Report by Preskill, H., Gopal, S., Mack, K. and Cook, J.

UNODC is an office within a wider system of organizations, the United Nations. This system is literally a system to change “the system”. The United Nations contributes to solving societal problems related to peace, security and development. Not a single day passes in which in your capacity as managers, decision makers or evaluators requires engagement with and management of complex situations and problems.

What is complexity?

In its most simple form complexity is the phenomenon that emerges from a collection of interacting objects.⁵⁰ There is not a unified definition of complexity but there is certainty about the elements comprising it. Complexity contains non-linear relationships among elements, behaves in an emergent, adaptive and co-evolutionary fashion guided by feedback loops (knock-on effect), displaying a mixture of ordered and disordered behaviour that gives the impression that the system has a life on its own (e.g. the stock market, a traffic jam, etc.).

A project or a programme is not always designed and implemented in a linear way. Internal and external factors such as political changes at country level, or internal changes of policies or donor requirements can have a strong impact on the plans designed and being implemented. Thus, there will be a need to adapt the project to the specific context where it is being implemented.

The project itself is a collection of elements interacting with a specific societal problem occurring in a local context. The project aims at improving the problem or situation. However, many projects or programmes are conceived and implemented as linear solutions to problems that are dynamic and multidimensional (human trafficking, corruption or terrorism cannot be tackled in a linear fashion. Using linear methods to design, implement and evaluate these types of programmes and projects imposes strong limitations to their effectiveness and impact.

How do you apply complexity to evaluation and programmes/projects?

UNODC interventions vary widely in terms of their levels of intricacy and complexity. One intervention alone can have some simple aspects, some complicated aspects, and some complex aspects. These differences have profound implications for evaluations. Few interventions are simple and have all the ideal aspects: a clearly defined outcome, a single causal pathway, SMART indicators, robust monitoring data, and stable implementation process that would make an evaluation relatively straightforward.

⁵⁰Johnson, Neil: *Two's Company, Three is Complexity* (2007)



Complicated interventions have many components, such as programmes providing technical assistance in the form of workshops, material and advice to improve the border security in a country. They are complicated for different reasons. They may have long causal chains with many intermediate outcomes, outcomes that require multiple interventions, or they may involve multiple implementing agencies with different agendas. Although they may have many moving parts, the parts generally come together in predictable ways to produce a result. In such cases, evaluations need to be sufficiently sophisticated in order to capture all the parts and interactions, and be clear about the limitations of the evaluation.

Complex interventions, on the other hand, are those with elements that are not predictable, such as regional programmes covering many diverse and interlinked mandates of UNODC (drug prevention, fighting against corruption, terrorism, etc.). Features of complexity include having to respond to constantly changing environments, uncertain funding, and new opportunities and challenges. Where many different dynamics are at play, it is more difficult to develop a ToC or logframe that clearly shows the pathways of change and the intended results. In such cases, evaluators may require more flexible evaluation processes, need to be attentive to inter-relationships, and benefit from incorporating elements of evaluative and systems-thinking. For example, instead of focusing on “what happened and why”, evaluators should be attentive to the need to question the evidence and expand their focus to “what works, for whom, and under what conditions”. There is also likely to be the need to look beyond linear relationships to interrelationships and issues of non-linearity. The resources below will be helpful in this regard.

ADDITIONAL RESOURCES

Inspiration for this chapter came from a variety of sources, several of which are listed below.

American Evaluation Association (AEA): an international professional association of evaluators focused on improving practice and supporting practitioners. Among its resources are free virtual professional development opportunities, including tips of the day, brief 20-minute webinars designed to introduce new tools, techniques and strategies, and more in-depth webinars. www.eval.org

Better Evaluation: an international collaboration dedicated to improving the practice and theory of evaluation by sharing information about tools, methods and approaches. www.betterevaluation.org

DME for Peace: a global community of practitioners, evaluators, donors and academics who share best and emerging practices on how to design, monitor and evaluate peacebuilding programmes. Through greater collaboration and transparency, this group works to increase the effectiveness of the peacebuilding field. www.dmeforpeace.org

EvalCommunity: a site to access information on evaluation-related events, resources, jobs and tenders, as well as to post resumes and find evaluators by country and area of technical expertise. www.evalcommunity.com

EvalPartners: an international forum aimed at strengthening national evaluation capacities jointly founded by the International Organization for Cooperation in Evaluation (IOCE) and UNICEF. A range of training opportunities, tool-kits, manuals and other evaluation resources including an e-learning programme in development evaluation can be found on its website at www.evalpartners.org

International Initiative for Impact Evaluation (3ie): an international grant making NGO that promotes and funds impact evaluation. Its site includes impact studies, research grants, practitioner tools, publications, events and other resources, including an expert roster. www.3ieimpact.org

Wageningen Center for Development Innovation, Wageningen University: a resource for approaching and applying theory of change. www.theoryofchange.nl

LISTSERVS

Eval Gender+ Website: a global partnership to promote the demand, supply and use of Equity Focused and Gender Responsive Evaluations under the Eval Partners umbrella. <http://evalpartners.org/evalgender>

EVALTALK: the official discussion list of the American Evaluation Association. <https://listserv.ua.edu/archives/evaltalk.html>

Monitoring and Evaluation News: a news service focused on developments in monitoring and evaluation methods. www.mande.co.uk

Pelican: a platform for evidence-based learning and communication for social change. Members post and answer evaluation-related questions and share resources. Join at: www.dgroups.org

PUBLICATIONS

Bamberger, M., J. Rugh, and L. Mabry, (2006). *Real World Evaluation*, Sage Publications.

Baptist, C. and B. Befani (June 2015) "Qualitative Comparative Analysis: A Rigorous Qualitative Method for Assessing Impact." *BetterEvaluation*. Retrieved March 2017 from: http://www.betterevaluation.org/en/resources/guide/qcr_a_rigorous_qualitative_method_for_assessing_impact



Corbin, J. and Strauss A. (2015). *Basics of qualitative research: techniques and procedures for developing grounded theory*. 4th edition. Sage.

Davies R., Dart J., *The most significant change (MSC) technique: A guide to its use*, United Kingdom and Australia, April 2005, available online at: <http://www.mande.co.uk/docs/MSCGuide.pdf>

Forss, K., M. Marra, R. Schwartz, editors (2011). *Evaluating the Complex: Attribution, Contribution, and Beyond in Comparative Policy Evaluation*, vol. 18, Transaction Publisher of Record in International Social Science.

Morra Imas, L. G. and Rist, R. C. (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations*. World Bank Group.

Patton, M. (2010). *Developmental evaluation applying complexity concepts to enhance innovation and use*. New York, NY: Guilford Press.

Squirrell, G. (2012). *Evaluation in action: theory and practice for effective evaluation*. Russell House Publishing.

Vogt, W. P., Gardner, D. C. and Haeffele, L. M. (2012). *When to use what research design*. Guilford Press.



UNODC

United Nations Office on Drugs and Crime

Vienna International Centre, P.O. Box 500, 1400 Vienna, Austria
Tel.: (+43-1) 26060-0, Fax: (+43-1) 26060-3389, www.unodc.org