SETTING UP COUNTERFACTUAL IMPACT EVALUATIONS (CIEs) FOR EU DEVELOPMENT COOPERATION

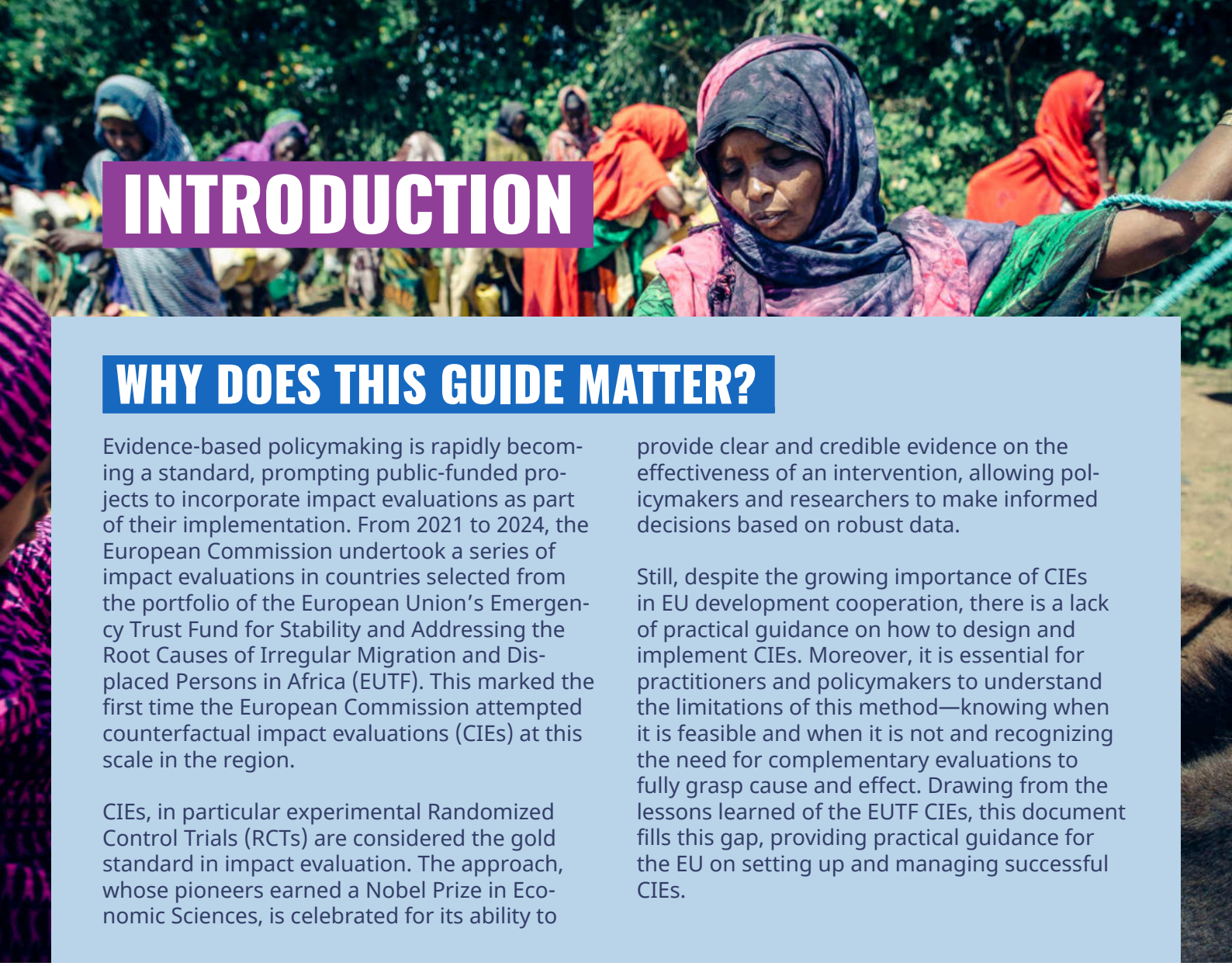
A PRACTICAL GUIDE FOR EVALUATION MANAGERS

Authors: Dr. Thomas Eekhout, Johanna Kern | February 2025

Center for Evaluation and Development

EUTF for Africa

EUROPEAN UNION



INTRODUCTION

WHY DOES THIS GUIDE MATTER?

Evidence-based policymaking is rapidly becoming a standard, prompting public-funded projects to incorporate impact evaluations as part of their implementation. From 2021 to 2024, the European Commission undertook a series of impact evaluations in countries selected from the portfolio of the European Union's Emergency Trust Fund for Stability and Addressing the Root Causes of Irregular Migration and Displaced Persons in Africa (EUTF). This marked the first time the European Commission attempted counterfactual impact evaluations (CIEs) at this scale in the region.

CIEs, in particular experimental Randomized Control Trials (RCTs) are considered the gold standard in impact evaluation. The approach, whose pioneers earned a Nobel Prize in Economic Sciences, is celebrated for its ability to provide clear and credible evidence on the effectiveness of an intervention, allowing policymakers and researchers to make informed decisions based on robust data.

Still, despite the growing importance of CIEs in EU development cooperation, there is a lack of practical guidance on how to design and implement CIEs. Moreover, it is essential for practitioners and policymakers to understand the limitations of this method—knowing when it is feasible and when it is not and recognizing the need for complementary evaluations to fully grasp cause and effect. Drawing from the lessons learned of the EUTF CIEs, this document fills this gap, providing practical guidance for the EU on setting up and managing successful CIEs.

FOR WHOM IS THIS GUIDE?

The guide is meant to inform EU evaluation managers & evaluation practitioners. In particular, this applies for the following departments:

- ► Directorate-General for International Partnerships
- ► Directorate-General for Neighbourhood & Enlargement Negotiations
- ► Service for Foreign Policy Instruments

WHAT IS THIS GUIDE ABOUT?

The guide provides a comprehensive and user-friendly resource that can be used by evaluation managers, implementing partners, and evaluation experts to improve their understanding and application of CIEs. It describes:

- ► Key concepts of CIEs.
- ► Key requirements to set up a CIE.
- ► Concrete guidance for evaluation managers on how to plan, set up and backstop a CIE in co-operation with evaluation experts and implementing partners.
- ► Practical tips and lessons learned based on the experiences from the EUTF CIEs.

What this guide does not provide is a step-by-step guidance on technical tasks that are conducted by the evaluation experts.

I. WHAT ARE CIEs AND WHEN SHOULD YOU USE THEM?

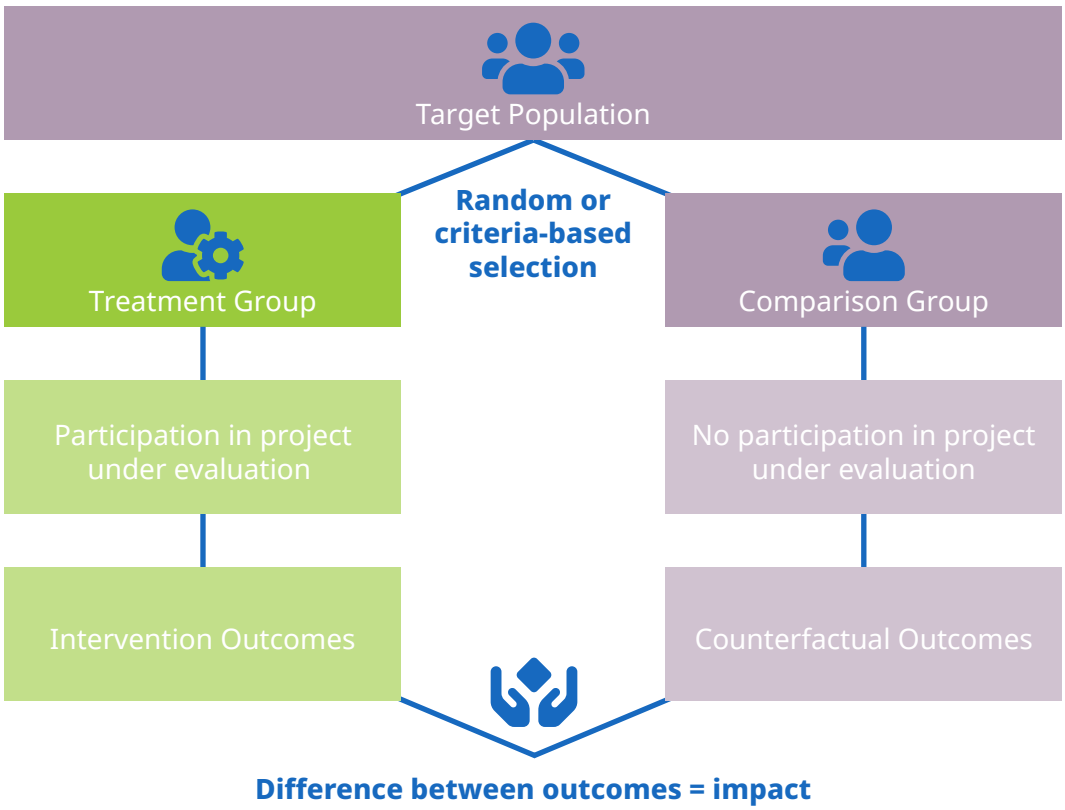
THE CONCEPT

The goal of a CIE is to measure the changes that can be attributed to an intervention. While conventional theory-based impact evaluations look at how and why an intervention is expected to work, CIEs are quantitative assessments that use a **counterfactual** to simulate what would have happened without the intervention.

The impact is essentially the difference between the actual outcomes and the counterfactual scenario. Having this counterfactual implies having data on the group of individuals, communities, or entities that receive the intervention - the "**treatment group**" - and a group that does not receive the intervention but is otherwise similar to the treatment group - the "**control group**" or "**comparison group**".

DIFFERENCES IN OUTCOMES BEFORE AND AFTER AN INTERVENTION ARE NOT "IMPACTS"...

... it is rather an assessment of changes over time which cannot be attributed only to the intervention because other factors can influence the same outcome. For example, if beneficiaries have a larger income after having participated in a training, it can be due to the intervention but also to an economic boom or other factors. A **CIE** analyses the causal impact of an intervention **by isolating the changes** caused only by the intervention.



WHEN ARE CIEs RELEVANT?

Before you decide to evaluate your intervention through a CIE you need to assess whether this **approach fits** the **objective** and **scope** of your envisaged evaluation. If the CIE is not fully fit for purpose, you should consider using other theory-based impact evaluation (TBIE) approaches (such as contribution analysis, process-tracing, outcome harvesting etc.) – which heavily feature qualitative methods – or mixed method approaches that combine CIE and other (qualitative and quantitative) methodological approaches.

You should choose a...

Counterfactual Impact Evaluation (CIE)

- ► To measure effectiveness / impact.
- ► To quantify the success of specific outcomes of an intervention in a clear and simple manner (e.g., "The [intervention] increased beneficiaries' income by 20% compared to the comparison group.").
- ► To compare the impact of various approaches, individually or in combination.
- ► To compare standardized measures across multiple interventions / countries.
- ► To understand what the impact would be on the overall targeted population.
- ► To understand what would happen if the intervention was scaled-up or implemented in other similar locations.
- ► To present clear and persuasive evidence to policy makers.
- ► If it is possible and ethical to establish a comparison group of non-beneficiaries.
- ► If topics are not too sensitive or complex to be explored through survey questions and to be quantified through (standardized) indicators or indexes.
- ► If the context is sufficiently stable and conducive for implementing the intervention as per the initial design.
- ► If a significant budget is available for the evaluation.

Theory-Based Impact Evaluation (TBIE)

- ► To understand what, how and why effects / impacts have materialized, or not.
- ► When the intervention design is likely to undergo future modifications.
- ► To test an intervention's theory of change and its underlying assumptions.
- ► When you want to understand the broader social, economic, and cultural context and provide a more holistic understanding of the intervention's impact.
- ► If stakeholder-engagement and participatory research are important aspects of your evaluation approach.
- ► When establishing a comparison group is impossible or unethical (e.g. humanitarian aid in emergency situation).
- ► When topics are sensitive or complex so that in-depth interviews are more useful than predefined structured questionnaires (e.g. intervention effects on intimate partner violence, gender norms, irregular migration intentions).
- ► You want to focus on the intervention's contribution (e.g. contribution of a component towards an overarching goal) and not necessarily on attributing its impact.

A Combination of Both ("Mixed Methods")

- ► When you have diverse evaluation objectives that go beyond one CIE or TBIE approach.
- ► When your evaluation is supposed to cover several of the OECD DAC criteria, including relevance, coherence, efficiency, and sustainability.
- ► When you also want to explore unintended effects.

DESIGN: HOW TO CREATE A COMPARISON GROUP?

The CIE design refers to how the evaluation is going to create the **comparison group**. The selection of the best-suited CIE design should be decided by an evaluation expert in consultation with the implementing partner, as it comes along with specific requirements and implications for the intervention. It is however important for the evaluation managers to understand that there are two families of CIE designs and their respective pros and cons.

EXPERIMENTAL DESIGNS (RCTs)

They are the golden standard to measure impacts. They rely on the fact that if the intervention randomly selects who will benefit from it before it takes place, those who do not benefit are similar, on average. To measure the impacts, one can simply measure the differences in the outcomes after the intervention.

- ► Most robust approach
- ► No obligation to collect baseline data
- ► Cheapest approach
- ► Straightforward-estimation models

- ► CIE must be planned before beneficiaries are selected
- ► Selection must be random
- ► Can be perceived as unethical

QUASI-EXPERIMENTAL DESIGNS

Not all interventions can randomly select the beneficiaries before they start. The non-random assignment means that participants are assigned to treatment or comparison groups based on pre-existing factors, such as location, eligibility criteria, or voluntary participation. In such cases, a counterfactual can be constructed. To measure the impacts one needs to use more complex models such as matching, difference-in-difference, or regression discontinuity approaches.

- ► More flexible approach
- ► Possibility to start CIE after start of intervention
- ► Often more feasible in real-world settings
- ► Useful when random assignment isn't ethical

- ► More complex estimation models
- ► Requires collecting baseline data
- ► Requires collecting data on more variables
- ► More expensive approach

Quasi-experimental methods and key concepts

► Matching: participants in the treatment group are paired with similar participants in the comparison group based on relevant characteristics (i.e. age, income, or education). So it tries to make the groups as similar as possible but isn't perfect since unmeasured differences might still exist.

► Difference-in-Difference: assumes both groups (treatment and comparison), even if they are different, would have followed the same trend if there were no intervention (parallel trends). It measures impacts by comparing changes in outcomes over time between groups.

► Regression Discontinuity: uses a cutoff point (like a test score) to assign participants to the treatment, and compares those just above and below the cutoff. It only compares individuals very close to the cutoff (e.g., a score of 49 vs. 51). It assumes these individuals are almost identical, except for the treatment. So if there is a sudden jump in outcomes right at the cutoff, it is attributed to the treatment.

BASELINE IS IMPORTANT, THOUGH NOT ALWAYS ESSENTIAL!

In a experimental design, conducting a comprehensive baseline is not essential. A short application form that collects information on key characteristics is usually enough (recall that in an RCT, selected & non-selected candidates are similar, on average). When conducting a quasi-experimental design, the baseline is essential since the evaluation experts need to measure and consider the differences between the selected candidates. Quasi-experimental designs are therefore more data and resource intensive than experimental designs.

EXPERIENCE FROM EUTF EVALUATIONS

Randomly selecting beneficiaries can be (perceived as) unethical. The buy-in of the implementing partners, political leaders and general population is necessary to conduct an RCT but some can be sceptical or uncomfortable at the idea of selecting who will benefit from an intervention by rolling the dice. For instance, randomly selecting who will be given money in a refugee camp can create conflictive situations between beneficiaries and non-beneficiaries.

The following steps utilized throughout EUTF can help to overcome the ethical concerns:

► Apply for and obtain the ethical clearance from relevant authorities.

► Communicate clearly to the intervention stakeholders the needs to undertake a RCT.

► By using a “phase-in RCT design” some randomly selected groups are treated before other groups, but all groups still receive treatment by the end of the intervention.

WHAT ARE KEY REQUIREMENTS TO FOR A CIE?

Once you have established that the CIE fits the purpose of your evaluation, there are several practical considerations that determine whether a CIE will be feasible and successful. Make sure the evaluation, the intervention under evaluation, its implementers, and evaluation experts fulfil the following **seven key requirements**.

1. HAVING A SUFFICIENT EVALUATION BUDGET

A CIE is relatively costly in comparison to other types of evaluations. Conventionally, the rule of thumb is that a CIE costs at least 200,000€ but depends on many factors such as the country in which the CIE is undertaken, the number of data collection rounds, how many aspects of the intervention are being evaluated, the characteristics of the sample and the use of complementary methods. See further details in section "defining the budget."

EXPERIENCE FROM EUTF EVALUATIONS

The average budget of an EUTF CIE was 376,384€, ranging from 290,313€ to 531,813€ in 2020. They often included three data collection rounds (baseline, midline and endline) to measure short- and long-term impacts, compared different trainings combinations to identify the most impactful approach and included qualitative interviews to better understand the impacts and investigate other criteria such as relevance, coherence and EU added value.

2. ANTICIPATING AN APPROPRIATE TIMELINE OF THE EVALUATION

Ideally, the planning of a CIE begins with the design of the intervention. It allows to consider the intervention, and evaluation needs before the start of the implementation. A CIE can start during or even after the intervention ended. However, you must consider that the range of CIE designs will be more limited.
The timeframe of a CIE depends on when the impacts are expected to materialise and on when the evaluation results are needed.

If a baseline is conducted, it needs to be done before the intervention starts.

EXPERIENCE FROM EUTF EVALUATIONS

All CIEs that started before the beginning of the intervention activities used an experimental design. Some of the EUTF CIEs began more than one year after the intervention started. It was therefore impossible to use an RCT since the selection of beneficiaries was not done randomly. Instead, the CIE experts used matching approaches. In one case, the CIE relied on baseline data collected by the implementing partner's team on all selected and rejected candidates. In another case, the CIE experts found a similar population outside the intervention's catchment to create a comparison group and collected baseline data retrospectively (however, collecting data retrospectively is not recommended since answers tend to be less precise and can be biased).
The EUTF CIEs generally lasted 4 years with data collections up to eighteen months after the end of the intervention as impacts tend to materialise on the long term.

3. VERIFYING THE UNDERSTANDING AND BUY-IN OF THE IMPLEMENTING PARTNER

It is crucial for the success of the evaluation that implementing partners understand and are committed to the CIE. For a smooth evaluation, the implementer must be particularly transparent and commit to the timeline of activities and the selection process.

EXPERIENCE FROM EUTF EVALUATIONS

In one RCT, the implementing partner agreed to randomly select beneficiaries from the sample for which baseline data had been collected. However, despite this agreement, the implementer failed to promptly inform the evaluation expert that the selection was not done randomly. This oversight had two significant consequences:

► It required collecting data on additional individuals who were not affected by the intervention's activities.

► The evaluation expert had to create a new counterfactual using a quasi-experimental design.

To increase awareness and understanding of CIEs among implementing partners and other stakeholders:

► The evaluation expert systematically gave an overview of the key concepts and their implications at the inception phase.

► EUTF invested in capacity building activities for EUTF partners on the key concepts, data collection activities and data analysis.

Linking capacity building with CIE activities facilitated a smooth implementation of evaluation activities and helped strengthen local capacities for future evidence-based programming.

4. SECURING A LARGE SAMPLE OF BENEFICIARIES

It is essential to have a large sample size so that a CIE has the sufficient "power" to detect an impact, if one has occurred. There is always a risk of not detect an impact with confidence even if it does exist. However, a large sample size reduces the risk of not detecting an existing impact. To determine the minimum/acceptable sample size, evaluators conduct power calculations.

EXPERIENCE FROM EUTF EVALUATIONS

The EUTF evaluation planned to conduct a CIE for nine selected interventions that initially met the requirements. However, two interventions unexpectedly each enrolled fewer than 100 beneficiaries due to a change in the interventions' strategy and limited applications. In such circumstances, a CIE would not have had the sufficient "power" to detect an impact. Instead, the two interventions were evaluated using only qualitative methods.

The CIEs conducted under EUTF used samples ranging from 1,200-3,300 beneficiaries.

5. BUILDING A GOOD COMPARISON GROUP

The comparison group is the cornerstone of the CIE as it represents the benchmark with which the beneficiary group is compared to in order to measure the impact of the intervention. It must be as similar as possible to the beneficiaries before the intervention starts and ideally remain unaffected by the intervention's activities during the duration of the evaluation. The counterfactual can be naturally produced by randomly selecting the beneficiaries or built using quasi-experimental methods. In case the comparison group is too affected, the evaluators will have to propose to switch to another evaluation design than the CIE initially envisaged.

EXPERIENCE FROM EUTF EVALUATIONS

A comparison group can be found in most cases. All EUTF CIEs could establish a comparison group. In the best-case scenarios, there were too many applicants for the intervention, and the comparison group was created by randomly selecting the beneficiaries and the non-beneficiaries. When a random selection was not possible, the CIE relied on other approaches. For example, some CIEs used rejected candidates after being ranked during the selection process. In other cases, the CIE sampled individuals living in the intervention's catchment area and that had similar characteristics as the beneficiaries.

6. ENSURING AVAILABILITY OF BENEFICIARY DATA

It is crucial that the implementing partner is willing and able to provide relevant data on beneficiaries. It is important to understand who benefits from what. Without monitoring of activities, one cannot know how implementation is evolving, track if objectives are being met or confirm an intervention's impact (or the absence of it) through evaluations. In other words, without monitoring data, the intervention remains a black box.

EXPERIENCE FROM EUTF EVALUATIONS

The CIE experts assessed the monitoring systems of the interventions evaluated to ensure it is useful for the CIE (i.e. monitoring system uses unique identifiers for each beneficiary and collects information on intervention participation). Since this was not the case in most interventions, the CIE experts supported the implementing partner in developing a useful monitoring system when possible.

Eventually, one CIE could not confirm whether the selected candidates received the support or not which raised challenges in the interpretation of the results: in the absence of a well-designed monitoring system, the CIE could not confirm whether the lack of impact was because the selected candidates did not benefit from the intervention or because the intervention did not adequately respond to their needs.

Consequently, the CIE findings were less reliable and conclusive as they had to be based on qualitative interviews and survey data only.

7. ENGAGING EXPERTISE IN CIE AND MIXED-METHODS APPROACHES

The evaluation expert responsible for the implementation of the CIE must have the know-how to plan and set up the CIE, implement large-scale surveys and analyse the data. However, a CIE expertise is not always enough. it is indispensable that the CIE expert:

- Keeps in mind that the principal goal is to assess the intervention's effectiveness and improve future interventions.
- Is familiar with alternative evaluations to CIEs in case the initial evaluation design is not feasible.
- Keeps in mind the implementation constraints to which IPs are confronted to.

EXPERIENCE FROM EUTF EVALUATIONS

The evaluations of the EUTF, which covered multiple countries and varied design elements (such as beneficiary selection, support mechanisms, and implementation timelines), required a high level of expertise in CIE. It was important to adapt methodologies to the specific needs of each intervention while upholding global standards for CIEs and meeting intervention objectives. This process also involved working with large and complex datasets, integrating mixed-methods findings, and presenting findings to both technical and non-technical audiences. A key takeaway was the value of contracting a research institute with experience in CIEs (matching, difference-in-difference, regression discontinuity...), mixed-methods evaluations, and in-house expertise. Additionally, using the same firm for all evaluations helped streamline coordination, ensure methodological consistency, oversee data collection with local partners, and maintain continuity throughout the 4-year period, despite challenges such as staff turnover and varying local contexts.

II. PLANNING THE CIEs

TAKING THE RIGHT STEPS IS VITAL

During the planning phase of the evaluation the donor and evaluation manager must define what must be evaluated, set a realistic timeline and budget for the evaluation and determine which and how stakeholders should be involved.

What & how do you want to evaluate?

When will the CIE take place?

How much will the CIE cost?

Who should be involved?

DEFINING WHAT TO EVALUATE

A first step in the evaluation process is to define the **use** and **scope** of the evaluation. These elements are usually set in the Terms of Reference of the evaluation, which is implemented by an evaluation expert (or team of experts), that is independent of the intervention. Think about **who** is supposed to use the evaluation findings **for what** and **why**. Consider that evaluations often serve different audiences and purposes.

To further define what impacts you want to evaluate start with **reviewing the intervention logic** of your intervention, which is often depicted through a Theory of Change (ToC). The development of a ToC is usually led by the implementing partner when conceptualising an intervention. It is key to map out what the intervention is expected to influence and how. Outlining the logic (including risks and assumptions) for the intended changes to occur, not only helps implementing partners anticipate mitigation strategies but also helps implementers and evaluation experts explore why impacts materialize or not. Take into consideration that stating unrealistic intended outcomes will probably lead the CIE to conclude that the intervention did not have any impact.

CIEs test whether expected outcomes of an intervention logic materialized or not. Choose those outcomes and causal chains that you want to verify and determine the scope of the evaluation by drafting a limited number of evaluation questions. An impact-related evaluation question usually has a simple structure:

"To what extend did [the intervention / activity] impact [the outcome]?"

This example illustrates that one must accurately define what activities are being considered and how the outcome will be measured. Evaluation experts often opt to use several indicators for each outcome. Note that CIEs usually cover sets of activities on which one can measure impacts and rarely cover the entire intervention (see Key requirements). If the objective of your evaluation is to gain other insights beyond attributing impact (e.g. if you want to contextualize or explain impact, assess efficient resource utilisation, implementation fidelity, or, (potential) sustainability of outputs or outcomes) other evaluation questions need to be added.

Once the objective and scope of your evaluation are determined, you need to determine the specific methodology of your evaluation. Most donors provide a broad framework for the methodology in their Terms of References and expect the evaluation expert to propose specific methodological approaches they deem most appropriate to answer evaluation questions. This has the advantage that you can rely on the expertise of the evaluation expert to determine the most suitable methodology while at the same time providing an overall structure for important evaluation determinants as time and budget.

EXPERIENCE FROM EUTF EVALUATIONS

Evaluation experts can help set up or refine a ToC in case implementers have not created a detailed ToC (yet). For some CIEs the evaluation experts drafted a refined ToC during the inception phase that helped inform the evaluation design.
All CIEs explored causal pathways at different levels of the results chain, including immediate and medium-term outcomes and long-term impacts. This provided a holistic understanding of the interventions and made it easier to identify where (and why) effects materialized as planned or not.

Most CIEs focused on specific components of the interventions as some activities did not comply with the requirements to undertake a CIE (mainly because they did not support enough beneficiaries or because their assessment would not have provided relevant insights for the evaluation).

All CIEs had a broader scope and applied different forms of CIE+ (mixed-method) approaches that used qualitative evaluation methods to explore how and why impacts materialised (or not) and secondary intervention data to assess cost-efficiency and implementation fidelity.

Terms of Reference provided evaluation experts with a methodological framework by specifying that a CIE should be conducted and that combined quantitative and qualitative methods should be used and left the development of country-specific evaluation methodologies to the evaluation expert.

SETTING THE TIMELINE

When the CIE should start and end are crucial aspects of the evaluation that need to be set in the evaluation's Terms of Reference. Ideally, the CIE should be planned with the design of the intervention. It allows to have a fully embedded design that incorporates the implications of the CIE. It also ensures that the implementing partner considers the needs of the evaluation expert and vice versa. RCTs can only be set up in this scenario as the implementing partner must agree to select its beneficiaries randomly before the start of the activities.

The CIE can also start after the design of the intervention. However, this can come along with complications:

- Absence of baseline data,
- Contractual obligations of the evaluation experts and implementing partner might not be fully aligned,
- The implementing partner may refuse requests from the evaluation experts if the CIE requires adaptations of the intervention design or the investment of additional resources.

When the CIE should end depends on when the impacts are expected to materialise and on when the results are needed. For some indicators it can take a long time before one can observe changes. For example, if you are investigating the impacts of a training on employment, impacts on beneficiaries are usually visible two years after the end of the intervention. If the CIE was designed after the intervention start its duration will naturally be shorter, but it rarely lasts less than 8 months. In some cases, the donor has a specific agenda that requires concrete answers from the CIE (e.g. development of a strategy, presentation at a conference). It can be helpful to identify the deadline for the CIE first and then identify possible methodologies and plan different steps accordingly.

PRINCIPAL STEPS OF A CIE

Intervention Activities

CIE Activities

- Setting up of intervention: Theory of Change, build monitoring system…
  Inception (Evaluation questions, indicators, evaluation design, planning of activities…)
- Identify eligible population (targeting, sensitization, call for applications)
- Selection of beneficiaries + Selection of comparison group
- **Baseline data collection**
- Implementation + monitoring
- **Follow-up data collection(s)**
  Timing to be adapted to the outcomes of interest
- Analysis and dissemination
- Incorporate lessons learned in next intervention

DEFINING THE BUDGET

CIEs are relatively expensive in comparison to other types of evaluations. The cost varies usually between 100,000€ and 500,000€. Most of the costs are due to the time dedicated by

- the evaluation specialist (including principal investigator(s),
- data manager(s),
- – if applicable – qualitative researcher(s),
- as well as the data collection activities.

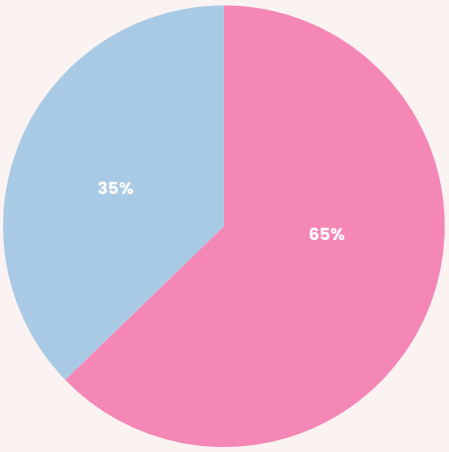
Although evaluation costs tend to exclude time and effort on the side of the implementing partner, donors should be aware that their involvement is needed for the facilitation of sampling and data collection, review and verification of research design and evaluation findings. The table below illustrates the approximative amount allocated to a CIE in the EUTF experience and its distribution across its different phases.

Experimental design are usually cheaper for two main reasons:

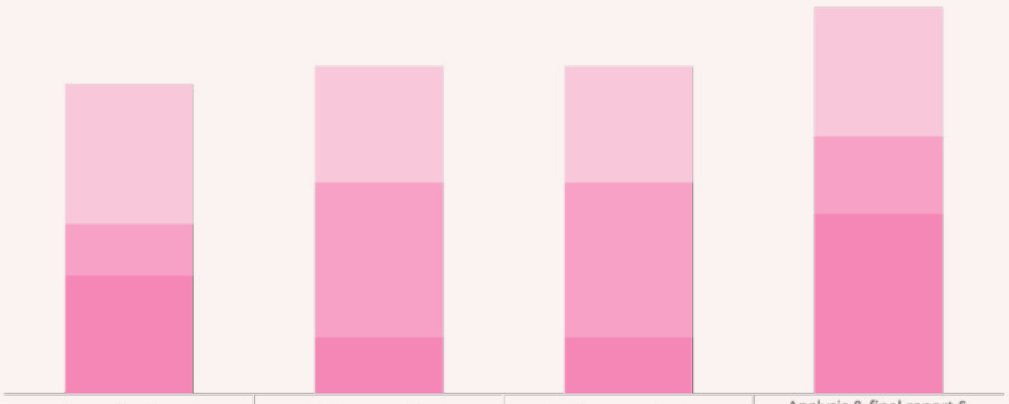
- An experimental design does not always require a (extensive) baseline. Collecting data on a few key variables is often enough. In fact, many experimental CIEs use data from the registration phase instead of conducting an independent baseline survey.
- A quasi-experimental design needs a larger sample size to be able to detect the same level of impact on an outcome than a experimental design. A rule of thumb is that quasi-experimental designs require a 20% larger sample as respondents in the comparison group might be deemed as irrelevant for the CIE (e.g. not comparable to the treatment group).

SAMPLE OF BUDGET DISTRIBUTION BASED ON EUTF CIEs

Data collection firm: One major cost is the cost of collecting data. Usually conducted by a local firm, based on the input and supervision of the CIE experts.



CIE Experts: The majority of spending is on experts, who guide the process, analyze the findings, and draw the conclusions on the impact of the intervention.



	Inception Phase	Baseline data collection	Endline data collection	Analysis & final report & dissemination
CIE: Qualitative researcher	18-28 days	8-12 days	8-12 days	28-42 days
CIE: Data Manager	8-18 days	24-36 days	24-36 days	12-18 days
CIE: Principal Investigator	14-20 days	20-30 days	20-30 days	20-30 days

EVALUATION EXPERTS' COST

The most expensive profile in the team of the evaluation expert is the principal investigator. He/she is essential to:

- design the best-suited evaluation plan,
- guide the design of the questionnaire,
- perform the advanced statistical analysis,
- write the evaluation report(s),
- oversee all evaluation activities.

A principal investigator often has rates ranging between 800€ and 1,500€ per day.

The data manager's tasks are primarily concentrated around the data collection activities. Specifically the data manager is responsible for the:

- design of the questionnaire,
- supervise the data collection and checking for the data quality,
- clean data,
- perform basic statistical analysis (e.g. description of the sample).

Rates of a data manager usually varies between 600€ and 800€ per day.

SUPPORT YOUR EXPERTS TO LIMIT COSTS!

To limit costs, hiring a data manager in addition to the IE expert is usually useful to perform the less methodological tasks related to data collection. To **maximise the quality** of the evaluation, the IE specialist should supervise all the evaluation activities.

DATA COLLECTION COST

It is rare that a CIE does not need to collect primary data and evaluation experts usually rely on a data collection firm based in the same country where the intervention is being implemented.

BE SMART TO REDUCE YOUR DATA COLLECTION COSTS!

Use the registration/application process of the intervention to collect data on all the candidates. This eliminates the costs of a baseline survey but requires a close collaboration between the implementing partner and the CIE to collect the relevant data for the evaluation and the intervention itself.

Collect data remotely. Nowadays, interviewees can often be reached by phone, sometimes even in remote areas. Conducting a survey via phones can reduce costs by more than half. The main drawbacks are that the interview should not last more than 30 minutes, enumerators can only rely on the respondent's answers and cannot verify their truthfulness.

The main determinants of data collection costs are:

► The country where data must be collected: this will determine the daily rates of the data collection team.

► The location of the interviewees: the more they are near to each other and accessible, the more interviews the team can do in one day and the lower the transportation costs will be.

► The number of the interviewees: a higher number of interviewees automatically increases the number of days that the data collection team must work.

► The length of the survey interviews: from our experience in Sub-Saharan Africa one interview costs 26€ on average and ranges from 11€ to 40€, covering the hiring and training of the data collection team, and conducting the survey.

IDENTIFY THE ACTORS OF THE CIE

The CIE team must be composed of the focal points from the different stakeholders involved in the exercise including the donor/evaluation manager, the implementing partner's focal point (usually a M&E officer) and a principal investigator of the evaluation expert. The data collection firm is also important, but it usually either part of the cost of the evaluation expert or sub-contracted by the evaluation expert. The table below lists the key domains of expertise and assets that the different team members should have to allow for a successful CIE.

Actors of the evaluation process		Domains of expertise / assets
Donor	► Evaluation manager (Reference group)	► Deep understanding of the intervention activities ► Decision power to adapt activities, if needed ► Evaluation coordination and quality assurance ► Decision power & understanding of CIE methods to approve evaluation deliverables
Implementing Partner	► M&E officer	► Familiar with evaluations, including CIEs ► Management of complex datasets ► At least basic statistical analysis ► (+ Has time available to support the CIE experts)
Evaluation experts	► Principal Investigator	► CIE expert ► Advanced descriptive and econometric analysis ► Knowledge in non-CIE evaluation methodologies ► Evaluation management ► Thematic area ► Geographic area
	► Data manager	► Large survey management ► Advanced descriptive analysis
Data collection firm	► Project manager ► Field coordinator(s) ► Supervisors ► Enumerators	► Local expertise ► Knowledge on administrative process to get ethical clearance ► Pool of experienced supervisors and enumerators ► Respects and is aware of the country's data collection ethical requirements

III. MANAGING THE CIEs

ENSURING OPERATIONAL SUCCESS

Once the donor/evaluation manager has answered to the key questions to plan the CIE, activities tend to become more technical. Implementation of activities then mainly fall on the evaluation expert who must ensure that the evaluation questions are answered with the agreed-upon evaluation method and within the set budget and timeline. The role of the donor/evaluation manager is principally to provide guidance, facilitation and quality assurance for the evaluation team, process and deliverables. The evaluation manager needs to backstop the evaluation expert, ensure that the CIE activities take place as planned, facilitate a smooth coordination between the implementing partner and the evaluation experts and ensure that evaluation process and outcomes adhere to quality standards. This section points out key elements to which the donor/evaluation manager should pay attention.

1. INCEPTION PHASE

For the evaluation expert, the first step is to define the best-suited CIE design within the parameters of the donor set in the Terms of Reference. This is done during the inception phase of the evaluation. During the tender process, the evaluation experts usually propose a preliminary CIE design based on the available information. However, the intervention might have changed, progressed or key information was missing which can lead to changes in the final CIE design. Similarly, the implementing partner is likely to suggest adjustments to the preliminary evaluation matrix developed by the intervention team. For these refinements, the evaluation expert may travel to the intervention site to get familiar with the context and the intervention team. It is also a key phase during which the evaluation and intervention team agree on roles, commitments and procedures. The inception phase usually ends with the evaluation expert submitting an inception report that details the planned evaluation methodology and the donor/evaluation manager (and reference group) approving the report.
The approval of the donor /reference group is not the only approval that matters. Before undertaking the evaluation activities, an institutional review board (IRB) and/or other relevant institutions must assess whether the CIE

AGREEING EARLY TO CONDUCT A CIE CAN FACILITATE GOOD COLLABORATION!

One must keep in mind that the smooth implementation of the intervention should remain the priority. The following steps have proven to be particularly useful:

► Agree on random selection of beneficiaries during the intervention design.

► Agree on collecting baseline data on all candidates including contact information before the intervention starts.

► Assign a unique identification number to each candidate during the registration and use it in all datasets.

► For each beneficiary, monitor the support received from the intervention.

If the CIE was not planned with the project, a Memorandum of Understanding can also promote the collaboration between the project implementer and the evaluation expert.

adheres to ethical principals. Obtaining this type of ethical clearance typically starts during the inception phase and needs to be concluded before data collection. Usually, it is the responsibility of the evaluation expert to obtain ethical clearance. Local presence of the donor, implementing partner or the data collection firm can help identify the responsible in-country IRBs. If there are no in-country authorities that provide ethical clearance, approval can also be given by international IRBs.

EXPERIENCE FROM EUTF EVALUATIONS

EUTF CIEs obtained ethical clearance by presenting the CIE design, tools and practical implications to the relevant institutions. The ethical review usually includes some form of assessment of risks to participants, safety and safeguarding of participants and researchers, informed consent, privacy and confidentiality, fair participant selection and compensation, scientific validity, conflicts of interest, and cultural sensitivity.

For EUTF evaluations requirements varied depending on the country/IRB and the nature of the study. In some cases, the ethical clearance was automatically obtained by collaborating with a certified data collection firm. In other cases, the ethical clearance was given by a national review board based in a university and for one study, by the Ministry of Education as the CIE delt with children. As there are no international standard procedures, the evaluation expert relied on the guidance of its in-country staff and subcontracted data collection firms.

EVALUATION MATRICES ARE THE BACKBONE OF A CIE!

Evaluation Matrices are often used as a planning tool to help organize and structure the evaluation process and findings. They ensure a systematic assessment of evaluation questions by clarifying through which indicators and data sources the evaluation questions will be answered. Survey data is the most common data source used for CIEs, although other sources are often particularly useful:

► It is recommended to complement survey data with primary qualitative data (from focus group discussions, in-depth interviews, key informant interviews, (semi-)structured and unstructured observations, and secondary data (particularly intervention planning and monitoring documentation)) as they often help better understanding the context of the CIE and how the impacts materialised (or why they did not materialise).

► An evaluation is often interested in understanding relevance, effectiveness, efficiency and coherence (and not only impact) which usually requires qualitative and secondary data.

2. DATA COLLECTION

In most cases the data collection is managed by the evaluation experts who must ensure that the data collected meets all the requirements of the CIE. Evaluation experts may have in-country personnel or hire a local data collection firm that is well established in the area and has the human and logistical capacity to rollout a large survey with experienced staff. However, in some cases, the donor may directly hire (and train) enumerators. In this case, the donor/evaluation manager must make sure the data is relevant, of high quality and can be used for the purposes of the evaluation. To ensure getting such type of data, attention needs to be given to the following:

► Those who collect the data answer to the research needs of the evaluation experts:

  ► The data collection firm collects the required information for the CIE: contact information, outcomes of interest (in the evaluation matrix), key characteristics (such as age, gender, household characteristics…).

  ► The roles, responsibilities and operating procedures between the data collection firm and the evaluation expert are agreed upon.

► The necessary steps are taken to collect high-quality data. This includes:

  ► Using a digital questionnaire that allows to upload data daily

  ► Testing the questionnaire

  ► Conducting a supervisor and enumerator training (usually ≥3 days)

  ► Supervising a pilot

  ► Undertaking daily checks of the uploaded data during the survey (i.e. High Frequency Checks)

► The evaluation experts check for the follow-up data collections if the timing allows to detect impacts ("is it too early or too late to see impacts?") and fits the institutional agenda ("will the CIE provide answers in time?").

MAKE SURE THAT THE EVALUATION EXPERT & INTERVENTION IMPLEMENTER COMMUNICATE!

The CIE cannot be efficiently undertaken if the project implementers and evaluation experts do not actively communicate with each other throughout the evaluation. The evaluation expert needs to understand challenges that the project implementers face and their consequences on how and when the activities are implemented since it can have implications for the evaluation. It is key that communication is timely and transparent on the following topics:

► How beneficiaries are selected.

► If selected candidates do not participate to activities or drop out.

► What and when activities take place.

The implementing partner can play an important role in this phase since it can provide primary data on beneficiaries, activities and outputs, which can be used to inform the evaluation. Collecting data on the candidates is a particularly useful initiative to set up the monitoring dataset and follow the support activities provided to each beneficiary but can also represent a great opportunity to save costs on an entire survey for the CIE. To build on such synergy, the evaluation expert and the implementing partner must agree before the survey on how and what data is collected.

3. ANALYSIS AND REPORTING

Apart from the inception report which defines the evaluation methodology as a whole, there are usually two other types of evaluation reports that the donor/evaluation manager can ask the evaluation experts to produce: the baseline report, and final report. The table below provides guidance to evaluation managers on the purpose, timing and structure of the reports.

The process of analysis depends on the types of evaluation methodologies used and often includes quantitative data cleaning, selection and estimation of the econometric model, and interpretation of the findings, the coding and analysis of qualitative data (through the development and refinement of themes and interpretation of data) as well as triangulation and validation of findings. While the analysis is typically led by evaluation experts, the donor/evaluation manager usually leads the validation, quality assurance and the dissemination of findings. They coordinate feedback to preliminary findings and draft reports of the reference group or other stakeholders involved in the quality assurance process. They also usually approve evaluation reports and decide how findings are further disseminated.

	Inception Report	Baseline Report	Impact Evaluation Report
When?	► Ideally, before the start of the intervention activities. ► At the latest, before the first data collection.	► After the baseline data is collected.	► After the endline data is collected. ► Before lessons learnt and recommendations are needed.
Why?	► Transparently communicate on a predefined research protocol. ► Agree with stakeholders (donor, implementing partner and CIE team) on the goals and CIE approach. ► Ensure ethical integrity and support request for ethical clearance. ► Enable replication and comparability. ► Anticipate potential biases and challenges. ► Provide a basis for adaptive planning.	► Establish pre-intervention conditions. ► Facilitate accurate impact measurement. ► Adjust evaluation design (if needed).	► Inform on what worked, why and to what extent. ► Inform on the challenges encountered by the beneficiaries and stakeholders. ► Summarise the main findings from the study. ► Share lessons learnt and recommendations for future interventions.
Outline	1. Introduction 2. Intervention logic 3. Stakeholder map 4. Evaluation methodology 5. Analysis of risks 6. Ethic rules 7. Evaluation questions 8. Work Plan 9. Annexes	1. Introduction 2. Intervention logic 3. Stakeholder map 4. Evaluation methodology 5. Data collection 6. Baseline findings 7. Implications for the CIE 8. Conclusions & Recommendations 9. Annexes	1. Executive summary 2. Introduction 3. Findings 4. Conclusions, recommendations & lessons learned 5. Annexes

The table below outlines roles & responsibilities across a CIE. Additionally, **approximate levels of effort** of the donor, implementing partner, evaluation expert and data collection firm throughout the different phases of the CIE are indicated by the **size of the marker**.

Phases	Activities	Stakeholders			
		Donor	Imple-menting Partner	Evaluation experts	Data collection firm
I. Planning of the CIEs	Define what to evaluate				
	Timeline				
	Budget				
	Set up the team				
II. Management & Coordination	Definition of IE design				
	Data collection				
	Coordination Eval. Expert & implementing partner				
	Analysis & Reporting				

 indicates the stakeholder is responsible

 indicates the stakeholder should be of support only

FURTHER READING

EU EVALUATION POLICIES AND GUIDES

European Commission: Directorate-General for International Cooperation and Development, **Evaluation matters – The evaluation policy for European union development co-operation**, Publications Office, 2014, data.europa.eu/doi/10.2841/85201.

European Commission: Directorate-General for International Partnerships, Hassnain, H., McHugh, K., Lorenzoni, M., Alvarez, V. et al., **Evaluation handbook**, Publications Office of the European Union, 2024, data.europa.eu/doi/10.2841/340793.

European Commission: Directorate-General for Budget, **Evaluating EU activities – A practical guide for the Commission services**, Publications Office, 2004, op.europa.eu/en/publication-detail/-/publication/6ff3c67d-bd1e-4909-8158-01cd57c4375d.

OECD DAC CRITERIA AND GUIDE

OECD (2021), **Applying Evaluation Criteria Thoughtfully**, OECD Publishing, Paris, doi.org/10.1787/543e84ed-en.

GUIDES ON PLANNING AND MANAGING COUNTERFACTUAL IMPACT EVALUATION

Czarnitzki, D. (2023). Guidebook for conducting counterfactual impact evaluations of State aid schemes for research development and innovation. Jrc External Study Report, 1-46, publications.jrc.ec.europa.eu/repository/bitstream/JRC135293/JRC135293_01.pdf.

Deiana, C. (2020). Counterfactual impact evaluation of European Social Fund interventions in practice, iris.unica.it/bitstream/11584/317438/1/EC_55495.pdf.

Naldini, A. & Kluve, J.& Pompili, M. (2021). Design and commissioning of counterfactual impact evaluations - Practical guide. 10.2767/02762. , www.researchgate.net/publication/357555350_Design_and_commissioning_of_counterfactual_impact_evaluations_-_Practical_guide.