

Guide for Terms of Reference

In this document we provide a guideline to write *Terms of Reference* (here-in-after ToR) in view of a counterfactual impact evaluation (CIE) project.

The ToR defines all the aspects of how an evaluation has to be properly conducted. **It presents an overview of the evaluation manager's requirements and expectations related to an evaluation study** providing a brief and concise description of the main scope and objectives of the evaluation, the roles and responsibilities of the involved actors, the methodology, the selection criteria, the timeline and the amount of resources available for the evaluation.

The ToR document typically performs **two distinct roles**: it serves as a step in the **procurement for evaluation services** setting the basis of a contractual arrangement between the commissioner of the evaluation (national or regional Managing Authorities (MAs) and an external evaluation consultant/team due to a competitive selection process. Secondly, it **sets the boundaries of the design of the evaluation** including the description of micro-data access/collection phases, the award criteria, and how to manage possible unforeseen hurdles.

A ToR should be structured so as to include the following basic elements:

1. a background *introduction and description of the intervention* providing context information, the main objectives of the program intervention (outcomes of intent) and the rationale for the evaluation;
2. the *specific scopes and objectives* of the evaluation identifying the main evaluation questions;
3. the *availability of access to relevant data* (e.g. outcome measures, covariates) on both the target population and the control group directly provided by the commissioning authority or publicly available;
4. the *expected methodological approach* eventually ensuring room for the evaluators to assess the quality of the proposed methodologies and eventually suggest additional ones;
5. a description of the *required professional competences and qualifications* of the evaluators (according to the scope and methodology of the evaluation) as well as of the *selection and award criteria*;
6. the *expected deliverables* (intermediate and final reports, presentations and so forth), the *time schedule* of the study and the available *budget*;

1. Introduction and Description of the Intervention

The first section of a ToR typically provides a general explanation of the intervention, program or project to be evaluated.

This section is meant to provide the evaluators with all the information about the purpose, objectives and intended outcomes of the program -including the expected\available outcome indicators-, **the rationale for the evaluation** (e.g. **why the intervention was chosen for evaluation**, what are we measuring, what change the intervention intended to induce on the outcome(s) of interest and how will the MAs use the results), a **description of the context** in which it was implemented, the roles and responsibility of the main stakeholders involved in the design and implementation of the program, information on the program itself and eventually of all the related studies, documents and evaluations that have been previously conducted.

What are the risks?

A too vague, too long or too confused introduction may be misleading for the evaluators. Specifically, if there are too many outcomes of interest, it is unlikely that the evaluation may address them all. Similarly, if the intervention is not well described, the reference population is not well defined and the design of evaluation will likely be incorrect. At the same time if other factors affect the outcome of interest, they may confound the effect of the policy intervention.

What are the remedies?

The introduction should be brief, precise and concise with a length of few paragraphs or a couple of pages depending on the complexity of the intervention.

In addition, the target population and the structure of the program intervention implemented have to be clearly described. Moreover, it is preferable to select a limited number of outcomes of interest preferring the ones that are **closer** to the intervention in a “theory of change” expected causal chain and to clearly define the reference population.

Selection Process of Participants

In a counterfactual evaluation design, the rough comparison between participants and non-participants in the intervention is potentially affected by *selection bias*. In order to tackle this problem, **it is necessary that a ToR provides the evaluators with an accurate description of the intervention implementation and the process adopted for the selection of participants, precis**ing what other factors in addition to participation are likely to affect the outcome of interest.

In particular, the following information needs to be specified:

- **eligibility criteria to apply for the intervention;**
- **criteria for the selection of participants among the applicants;**
- **what other factors in addition to participation are likely to affect the outcome.**

These details enable the evaluator to propose the best evaluation strategy for the assessment of the effects of the intervention.

2. Specific Scopes and Objectives

An important role of the ToR background introduction is to describe what are the main purposes, the expected effects and outcomes of interest of the intervention. Once the key objectives of the program have been introduced, the following section of a ToR needs to detail the specific evaluation questions identified for each program's objective presented. In addition, it is important to specify the underlying "theory of change" that links the intervention to the expected changes in the outcomes of interest. **Hence, a counterfactual evaluation is aimed to rigorously measure whether part of the observed changes in the available outcomes is caused by a program intervention and to desirably define the size of this effect. The statement of the specific objectives for the evaluation is crucial to define the main questions that the evaluators might be required to answer.**¹ For these reasons, it is indeed important to specify the availability of data regarding the outcomes of interest and other factors for the reference population.

What are the risks?

An unclear definition of the evaluation objectives might lead to a deliverable which will not correspond to the expectations of the MAs.

The lack of a "theory of change" renders the evaluation not credible while the lack of available data may render the evaluation impossible even in presence of a sound "theory of change".

A too long list of objectives might undermine the feasibility of the study.

¹ Please note that for an overview of the existing literature on the CIE of active labour market policies and programmes, you can visit the Counterfactual Evaluation Archive on CRIE website (https://crie.jrc.ec.europa.eu/CIE_database/cieDatabase.php).

What are the remedies?

It is important to clearly explain the **logical connection between the rationale for the evaluation, the “theory of change”, the declared intervention objectives, the evaluation purpose and the evaluation questions.**

For these reasons **the number of evaluation questions needs to be limited to 5 well defined questions at most.** It is generally preferable to accurately define few specific questions close to the output of the intervention in details rather than to ask the evaluator to examine a broad set of general questions.

The type and purpose of the counterfactual evaluation questions should be defined according to the type of program or activities evaluated.

In addition, **a prioritization of the assigned tasks along with a definition of the general evaluation assessment process** is crucial. It is important that this section **does not anticipate the evaluation results: the theory of change expresses the rationale for the intervention and the intended causal chain while the evaluation has the scope to say what the effect was in reality.**

3. Data Availability

The ToR should include a detailed description of the data that will be provided by the commissioning agency to the selected team of evaluators. The ToR should clearly anticipate where the data can be found and state the rules of the game for the contractors (e.g. specify if administrative data is available or if the contractor will be required to collect data itself through a survey).

What are the risks?

If no detailed provision of data is made, the contractor may not be able to carry out the required work.

The evaluation work could be delayed and the quality of the CIE could be hampered because of data issues related to availability and quality.

If data is provided to the contractor the cost of the contract will be much lower than otherwise and the conducted evaluation is usually of better quality.

What are the remedies?

Data sources: Describe the availability of the appropriate data sources (survey, administrative or both), their level of readiness-to-work and overall quality. This section

should also include a detailed description of all the available variables contained in the datasets, distinguishing among outcome(s), treatment and control variables.

Data should include **information on the program's recipients as well as on people who are not recipients of the intervention**. The latter might therefore be considered as a **control group** for the evaluation study.

Recipients (i.e. participants or treated) and non-recipients (i.e. non-participants or controls) have to belong to the same population the intervention is targeted at. This population, defined as **target population**, is identified through the eligibility criteria established for the specific policy or intervention.

The information on these criteria, as well as on the selection process of participants, is in fact fundamental to tackle the potential *selection bias* and compare properly the two groups of participants and non-participants in the intervention.

In the ToR the MAs should indicate what type of data the evaluators can access in order to fulfill the minimum requirements of the evaluations study. Analogously, the ToR should leave room for improvement depending on the possibility to access additional data.

In the best case scenario the MAs should guarantee the use and the access to **administrative data**, which can ensure the availability of all relevant variables and therefore allow better quality evaluations.

If administrative data are not available for the evaluation and the data sources provided for the CIE study do not include information on not participants, MAs have to envisage the use of **additional data sources** in order to provide the evaluators with this type of information.

Data quality: Regardless of the source, **data** collected would need to be **representative** for the target population. This means that everybody in the target population needs a non-zero chance of being collected. In particular, the **sample** of participants and non-participants need to be two representative sample of the same target population².

Data protection: If **confidentiality of data** is explicitly required due to the nature of the data sources involved, the ToR should **clearly specify the additional requirements on ethical and professional standards to be guaranteed** by the evaluators (such as specific data protection requirements or anonymisation procedures and/or the need of establishing a data protection manager/officer).

² Please note that for the definition of the samples size you can refer to the tool available on CRIE website (<https://crie.jrc.ec.europa.eu/sampleSize/samplesize.php>).

4. Methodological Approach

The available data allows for the application of a **range of the counterfactual methodologies** (e.g. a family of methods) for the evaluation plan that should be described by the MAs in a separate ToR section. The set of counterfactual methodologies shall be consistent with the evaluation questions and the available data.

What are the risks?

Not all methods can be applied in a given context. The contractor may adopt a trial and error approach in order to select the best methodology that gives results close to expectations. No sophisticated methods can compensate for poor data quality.

What are the remedies?

It is important to specify the category of methods (counterfactual) to be used in the evaluation and to not leave this choice to the evaluator.

Specifying the methodological framework adopted for the evaluation study should also guarantee to the MAs the adoption of an appropriate evaluation methodology by the evaluators that is in line with and can answer the prescribed ToR's scope and objectives. However, it should also ensure enough flexibility for the evaluators to modify the evaluation plan in their proposals in order to possibly improve the proposed methodologies. The evaluator should say why this improves the design providing supporting evidence that the assumptions underlying each method are valid and appropriate in the given context.

5. Selection and Award Criteria

In this section the criteria for the selection of the evaluator team among all the potential candidates have to be clearly specified.

The ToR should indicate **the minimum requirements on size and experience profile of the evaluation team**, the **qualifications of the principal investigator**, the **necessary human and technical resources**, and the **distribution of responsibilities** among team members to perform the proposed work and **demonstrated specific experience** usually from relevant evaluations performed in the past.

It is particularly important in this section to require the evaluators to provide some quality control arrangements throughout the evaluation process. The MAs should ask the bidders to produce a detailed timetable of the planned tasks and subtasks identifying potential risks and challenges linked to the evaluation assignment and detailing how they plan to manage and overcome those issues.

The selection criteria typically award bids on the basis of the quality and feasibility of the proposed methodological approach, the expertise and experience of the researchers' team and the offered price.

What are the risks?

Usually two award criteria are used, price and quality of the project. However, because of uniform quality evaluations in the selection procedure (e.g. little variations among projects) and/or budget constraints, the price criteria might become the effective award criteria. This usually results in poor evaluation teams awarding the contract.

What are the remedies?

It is particularly important that **the selection criteria are described with clarity and objectivity** and that they **really differentiate between bids avoiding too high weights on the tender costs**.

In order to assign proper weights also to quality criteria it is important to invest on the quality assessment of proposals. The selection panel should include CIE independent experts, in order to provide an in-depth evaluation of the methodological aspects of the proposals and to be able to differentiate on quality and not only in price.

6. List of Deliverables and Time Schedule

Finally, **it is necessary to specify what are the intermediates and final deliverables (usually inception, intermediate and final reports) and the corresponding timeline** for the study. More precisely, it is important a ToR requires several quality checks throughout the evaluation process (such as public presentation with discussants -e.g. CIE experts) asking the bidders to provide a detailed timetable of the tasks and subtasks, to describe their planned quality control arrangements, to identify potential risks or challenges linked to the evaluation assignment and to detail how they would manage and overcome those challenges. In addition, if a ToR requires the hand-in of both the evaluation study and the underlying data the replicability of the analysis is guaranteed.

What are the risks?

In the outsourcing procedure, the risk that the MAs commissioning the CIE study might encounter is that the delivered study does not reflect the required tasks and the expected quality. In particular, the evaluation project might go astray if it is not guided by mentors and discussants during the different phases of the project.

What are the remedies?

In this schedule it is important to provide **reasonable deadlines** clearly based on the amount of products required to the evaluators. Short turnaround time usually generates poor quality reports

while excessively long times with no checks are useless. The involvement of external experts may be useful for the assessment of deliverables.

To minimize these risks, **the intermediate deliverables (inception, intermediate and final reports) should not be too close over time (3-4 months each) in order to leave to the evaluator some room to manage her/his work.** It is particularly important to define the outputs and reporting deliverables including some buffer time in order to ensure the evaluation meets the expectations avoiding delays in completing the assigned work.

It is fundamental to define the milestones of the evaluation project and to use the intermediate deliverables to ensure a quality check over time, usually involving external experts (such as CIE experts and/or academic researchers) as discussants in all milestones deliverables.

If the involvement of external experts may be useful for the different work phases related to a ToR (writing, data procurement, quality check on the CIE study and in particular on the methodological aspects, etc.), DG EMPL and CRIE will also be happy to provide their support to the MAs.³ Therefore, do not hesitate to contact us for additional support you might need, using the different channels at your disposal (Yammer, email, CoP, CRIE website: <https://crie.jrc.ec.europa.eu/>).

³ Please note that as part of the initiatives to promote the use of CIE and facilitate the outsourcing of CIE of ESF funded interventions, DG EMPL and CRIE have recently launched the CIE Market on CRIE website (<https://crie.jrc.ec.europa.eu/?q=content/cie-market>).