

Impact Evaluation Guidance for Process and Content

Impact Evaluation Quality Assurance System



Foreword

The Impact Evaluation Quality Assurance System (IEQAS) is one of the building blocks for implementation of the 2022 WFP Evaluation Policy. As such, it is the evaluation function's primary means of safeguarding international evaluation principles of:

- **independence:** by setting standards that increase the impartiality of the evaluation process and reporting on findings;
- **credibility:** by setting standards that ensure evaluations are evidence-based and follow transparent and systematic processes; and
- **utility:** by building milestones into evaluation processes for timeliness and reporting standards to ensure accessibility and use.

The IEQAS guides all impact evaluations conducted by WFP. The IEQAS is a working tool for WFP's evaluation cadre, evaluation partners and evaluation teams, covering all stages of the impact evaluation cycle. It is not a comprehensive handbook on evaluation and does not replace the range of evaluation literature available.

The IEQAS builds on the norms and standards of the United Nations Evaluation Group (UNEG); the OECD-DAC Evaluation Network; related tools from the Active Learning Network for Accountability and Performance (ALNAP); and the wider evaluation literature and community of practice.

The IEQAS pack consists of:

- 1 guidance for process guide and content;
- 2 quality checklists;
- 3 templates;
- 4 technical notes;
- 5 post-hoc quality assurance template; and
- 6 Other reference materials.

In 2020, the Office of Evaluation (OEV) reviewed its entire Evaluation Quality Assurance System across all types of evaluations to ensure closer alignment where required and to reflect recent UNEG guidance, such as the 2020 *Ethical Guidelines for Evaluation*. The first draft of the IEQAS was developed in 2023. It will be periodically and systematically updated in line with the WFP evaluation function's evolving needs, international best practice and feedback from country offices, Regional Evaluation Units and partners to ensure that it continues to serve its intended purpose. Further updates and new material will continue to be added as needed, to ensure the IEQAS continues to reflect emergent best practice and management requirements.

Anne-Claire Luzot, Director of Evaluation, March 2024

Contents

Introduction.....	1
What is an impact evaluation?	2
What is an impact evaluation window at WFP?	4
Roles and responsibilities	5
Phase 1: Planning	8
1.1 Planning phase process guide	10
1.2 Planning phase outputs.....	11
1.3 Planning phase content guide and quality standards	11
1.4 Reference materials for planning phase	11
Phase 2: Preparation.....	12
2.1 Preparation phase process guide	13
2.2 Preparation phase outputs	14
2.3 Preparation phase contents guide and quality standards	14
2.4 Reference materials for preparation phase.....	15
2.5 Identification of the counterfactual	16
3.1 Process guide for identification of counterfactual phase	16
3.2 Identification phase outputs.....	17
3.3 Reference materials for randomization phase.....	17
2.6 Optional pilot impact evaluation (lean approach)	18
2.6 Pilot impact evaluation process guide	19
2.7 Content guide and quality standards for pilot impact evaluations	19
2.8 Reference materials for pilot impact evaluation	20
Phase 3. Inception	21
3.1 Inception phase process guide.....	22
3.2 Inception phase outputs.....	24
3.3 Inception phase contents guide and quality standards.....	24
3.4 Reference materials for inception phase	24
3.5 Baseline analysis (optional).....	25
Phase 4. Implementation monitoring and/or high-frequency data collection	29
4.1 Implementation monitoring.....	29
4.2 Midline analysis (optional).....	30
4.3 Midline analysis content guide and quality standards.....	30
4.4 Reference materials for midline analysis	31
Phase 5. Endline data collection	32
5.1 Endline data collection phase outputs	33
5.2 Endline data collection phase content guide and quality standards	33
5.3 Reference materials for endline data collection phase.....	33
Phase 6. Analysis and reporting	34
6.1 Final evaluation phase outputs	36
6.2 Final evaluation report phase contents guide and quality standards.....	36
6.3 Reference materials for the report phase.....	36
Phase 7. Dissemination and learning.....	37

7.1 Follow-up and dissemination phase process guide.....	37
7.2 Reference materials for the follow-up and dissemination phase.....	39
Annex A: Ethics in impact evaluations.....	40
Respect for persons.....	40
Beneficence	41
Justice	42
Annex B: Roles and responsibilities matrix.....	43
Acronyms	46

Introduction

The [WFP Evaluation Policy 2022](#) defines impact evaluations as measuring changes in development outcomes of interest for a target population that can be attributed to a specific programme or policy through a credible counterfactual. According to the [WFP Impact Evaluation Strategy \(2019-2026\)](#), WFP defines the counterfactual as estimating what would have happened in the absence of the intervention – or establishing that outcomes for the beneficiaries would not be present without the intervention.

Overview of the guidance: These guidance materials apply to the management and conduct of impact evaluations. They are structured following the eight phases of an impact evaluation, focusing on the processes, outputs and quality standards that will be used for each of them. The seven phases are:

- 1 planning;
- 2 preparation (including a pilot impact evaluation if/where appropriate);
- 3 inception (baseline analysis if/where appropriate);
- 4 implementation monitoring (midline analysis – where appropriate);
- 5 endline data collection;
- 6 analysis and reporting; and
- 7 dissemination and learning.

This **process guide** shows the roles and responsibilities of each stakeholder during an impact evaluation. The purpose of the process guide is to provide a step-by-step description of the process leading to a country's inclusion in an impact evaluation window (IEW), highlighting the roles and responsibilities of each stakeholder. As impact evaluations are delivered through a hybrid model, Impact Evaluation Teams (IETs) combine expertise from the Impact Evaluation Unit (IEU) and external partners. Each IET includes at minimum an evaluation manager (EM), IEW coordinator (WC), who is often the EM, evaluation analyst (EA), technical lead (TL) and field coordinator (FC).

As impact evaluations generally take place in a thematic IEW, this guide outlines how country offices engage the IEW and provides a road map to guide the implementation of an impact evaluation within WFP. It describes the governance structure, the evaluation phases and steps, the quality assurance process, the post-hoc quality assessment system, and ethical standards. Each section includes links to relevant supporting documents and templates.

The content guides and quality standards are provided for the outputs produced during each of the evaluation phases. This guidance provides a brief introduction to general principles. Templates and a quality checklist for each product are used by the EM, second-level quality assurer, evaluation analyst, technical lead and FC.

Links are provided to other Office of Evaluation guidance, such as **cross-cutting technical notes**.

Key **UNEG guidance** such as the *UNEG Norms and Standards*, the *UNEG 2020 Ethical Guidelines for Evaluation*, the UNEG guidance for integrating human rights and gender equality in evaluation, and the [United Nations Disability Inclusion Strategy](#) are also referenced where relevant.

Impact evaluation phases and timeline: The start of an impact evaluation in any country is preceded by significant lead time within the framework of the IEW. The total duration of an impact evaluation significantly varies depending on the duration of the intervention, expected timing for the outcomes to materialize, and whether a pilot phase is required or not. The estimated time taken for each phase is shown in Table 1 below.

Quality assurance: Quality assurance cuts across all phases as it includes quality assurance of the evaluation processes and clearance of all evaluation reports.

First-level quality assurance (QA1) is provided by the EM (typically the WC), who is part of the WFP Impact Evaluation Unit. The EM is responsible for ensuring the evaluation responds to the evidence priority needs, and the WC is responsible for ensuring the correct implementation of the evaluation within the window. S/he is supported by the evaluation analyst, the WC (if not already EM) and the Head of the Impact Evaluation Unit. The EM is responsible for managing the IET to ensure the correct implementation of all the evaluation phases defined in this process guide in coordination with the partners and stakeholders to ensure quality processes. Finally, the EM is responsible for ensuring the correct quality assurance process

for all the evaluation products ahead of their review by second-level quality assurance (QA2) and approval by the Director of Evaluation.

Table 1: Estimated timeline (in months)

Phase	Minimum	Maximum
1. Planning	1	12
2. Preparation (optional pilot phase adds 2-12 months)	1	18
3. Inception (Optional inception Note and/or baseline analysis where appropriate)	2	6
4. Implementation monitoring (midline analysis where appropriate)	2	48
5. Endline data collection	2	4
6. Reporting	6	12
7. Dissemination and learning	1	12

QA2 is provided by the Head of the Impact Evaluation Unit. Where appropriate, QA2 may directly provide a recommendation for clearance to the Director of Evaluation.

Final approval of reports. All final reports are approved by the Director of Evaluation before publication by WFP.

Post-hoc quality assurance: Final impact evaluation reports are assessed by independent external reviewers to indicate the overall quality of the evaluation. Post-hoc quality assurance identifies a set of eight criteria, each with a set of elements. Each element is independently scored, and then aggregated to generate an overall four-level rating for the evaluation report, which ranges from Highly satisfactory to Unsatisfactory. A summary of the assessment for each report is published externally, along with the final report, on the WFP website.

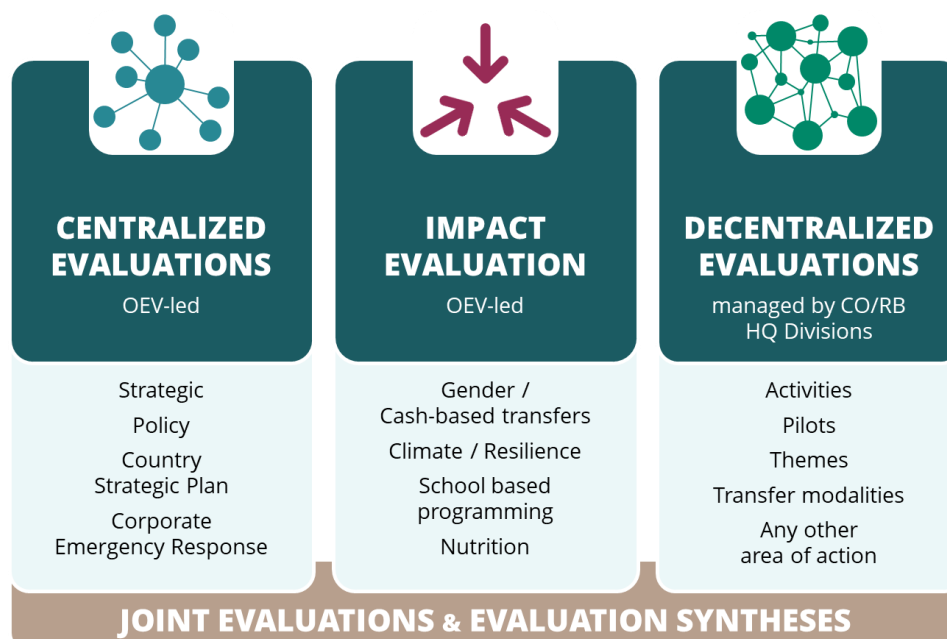
Ethics: Ethics are a key component in all impact evaluations. All impact evaluations must secure ethical approval from an internationally recognized Institutional Review Board (IRB) and, if applicable, from a national authority in the country where the impact evaluation is to be conducted. IET members who handle personally identifiable information must also have an up-to-date Protecting Human Research Participants certificate, which ensures that they are aware and knowledgeable of practices to protect participants in the evaluation. Annex I describes the principles of Respect for Persons, Beneficence and Justice, which will be used by the IRB during the review of the evaluation design, questionnaires and protocols. For more information, the [Ethical Guidelines for Implementing Impact Evaluations in WFP](#) outlines the considerations at each phase of implementation, their corresponding ethical commitments and other efforts practiced by the (IEU) of the Office of Evaluation (OEV) in order to maintain high ethical standards.

WHAT IS AN IMPACT EVALUATION?

The [WFP Evaluation Policy 2022](#) identifies three categories of evaluations. Centralized evaluations are commissioned and managed by the Office of Evaluation and conducted by external evaluators. Decentralized evaluations are commissioned and managed by country offices, regional bureaux or headquarters-based divisions other than the Office of Evaluation and conducted by external evaluators. Impact evaluations are managed by the Office of Evaluation following a request for support from a country office and feasibility assessment by the Impact Evaluation Unit.¹ Impact evaluations are included in windows following a recommendation from the Window Steering Committee (WSC) and subsequent approval by the Director of Evaluation. Impact evaluations are conducted by the Impact Evaluation Unit jointly with external technical partners (such as the World Bank’s Development Impact (DIME) department), and in close coordination with relevant WFP teams at headquarters, regional bureau and country office levels, and cooperating partners as appropriate.

¹ [Impact Evaluation Decision Guide](#).

Figure 1 Evaluation categories at WFP



The [WFP Evaluation Policy 2022](#) defines impact evaluation as “measuring changes in development outcomes of interest for a target population that can be attributed to a specific programme or policy through a credible counterfactual”. WFP defines the **counterfactual** as estimating **what would have happened in the absence of the intervention** – or establishing that outcomes for the beneficiaries would not be present without the intervention. WFP impact evaluations are prospective, meaning they are planned and designed prior to programme delivery or a new phase of intervention.² Impact evaluations align with the timeline of a programme or pilot and usually cover one or more years. Country offices play a key role in each impact evaluation by providing funding and ensuring impact evaluation designs stay aligned with programme implementation.

The term “impact” is understood as short-term or long-term “changes and effects”, and is not to be confused with impact indicators in a results chain. Impact evaluations are useful to demonstrate the impact of an intervention to support decision making, particularly to:

- assess the cost-effectiveness of new and innovative programmes;
- inform strategic decisions on whether to scale up innovations and pilots;
- test whether a programme is replicable in a new context; and
- test causal pathways and delivery mechanisms.

Due to feasibility (timing, sample size, budget, etc.), WFP impact evaluations usually cover only a limited set of interventions at a time, and attempt to answer only a limited, carefully selected set of questions. A significant amount of time and effort is usually required to arrive at the best impact evaluation design for a particular programme or pilot. This means that, for many interventions, it may not be feasible to design a rigorous impact evaluation. It is therefore important to consider impact evaluation as one tool in a wider spectrum of evidence-generating activities.³ Additionally, impact evaluations are primarily learning exercises, and WFP has decided not to establish a coverage norm for this type of evaluation. Individual impact evaluations do not produce recommendations or management responses and are not required to

² In exceptional cases where data are available, an impact evaluation may be retrospective.

³ The [Impact Evaluation Decision Guide](#) will help country offices in determining whether an impact evaluation is feasible and relevant for their interventions.

be presented to the WFP Executive Board.⁴ However, the Office of Evaluation does periodically update the Board on impact evaluation evidence generated by windows.

WHAT IS AN IMPACT EVALUATION WINDOW AT WFP?

The [WFP Impact Evaluation Strategy \(2019-2026\)](#) identifies four high-level objectives that guide all impact evaluation processes:

- **Objective 1: Contribute to the evidence base for achieving the Sustainable Development Goals (SDGs).** WFP uses its position as the world's largest humanitarian organization fighting hunger and malnutrition to ensure impact evaluation evidence contributes to global knowledge on what works best to achieve the SDGs.
- **Objective 2: Deliver operationally relevant and useful impact evaluations.** WFP uses robust impact evaluations to test programme theories, and learn what works best, how and for whom.
- **Objective 3: Maximize the responsiveness of impact evaluations to rapidly evolving contexts.** Impact evaluations are supported and delivered in a manner that responds to evolving contexts, maximizing opportunities to improve performance and optimize interventions.
- **Objective 4: Harness the best tools and technologies for impact evaluation.** WFP impact evaluations harness the best possible tools for capturing and analysing data to generate relevant insights into what works best in humanitarian and development interventions.

Impact evaluations are primarily delivered through thematic IEWs, in partnership with programme teams and co-funded by participating country offices. The windows are portfolios of impact evaluations, managed and co-funded by the Office of Evaluation, that generate evidence in WFP's priority areas.

The aim of the IEW is to stimulate and shape demand for impact evaluations in priority areas and enable the Office of Evaluation to prepare cross-regional portfolios that allow for the kinds of evidence syntheses that meet WFP's evidence needs. The Office of Evaluation gives priority to impact evaluations that fit within the thematic IEWs and funds the technical assistance portion of these evaluations. **Impact evaluations conducted within each window will therefore answer country-specific questions as well as window-relevant questions.**

IEWs are identified through a consultative and participatory process with headquarters and regional bureau technical staff, selected country offices, and research partners. The Office of Evaluation, in close collaboration with relevant technical units, conducts an in-depth review of existing evidence in an identified thematic area to identify evidence gaps. The Office of Evaluation organizes a participatory workshop, bringing together key stakeholders to present and discuss emerging questions and evidence needs that emerge through the literature review and during the consultation process.

Current windows include:

- 1 Cash-based Transfer and Gender Window;
- 2 Climate Change and Resilience Window; and
- 3 School-based Programmes Window.

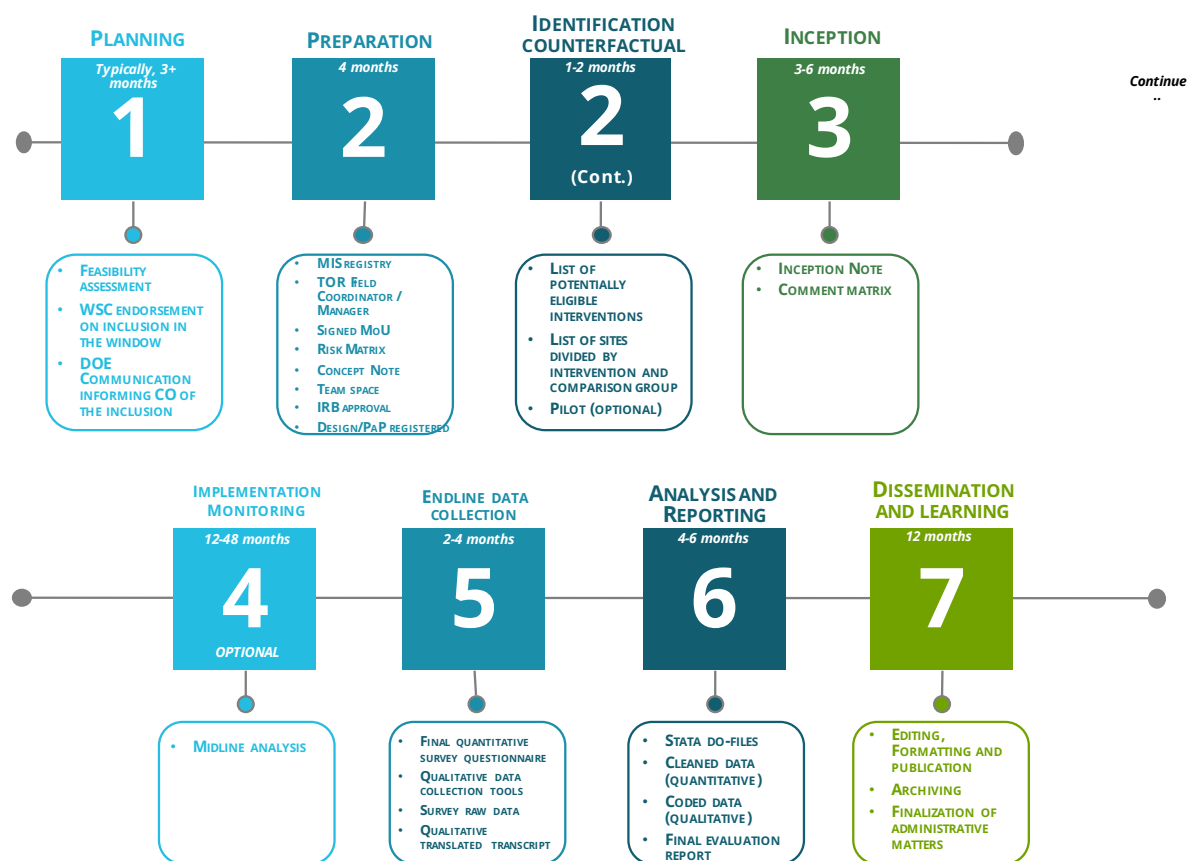
The Office of Evaluation announces calls for expressions of interest to country offices approximately once a year for each thematic window and can, on an exceptional basis, accept new proposals with particularly high relevance to a window between formal calls for expressions of interest. The likelihood of the Office of Evaluation taking on an impact evaluation is higher if its evaluation questions are relevant to one of the thematic windows. IEWs can stay open and update their evidence priorities over time to align with WFP's evolving evidence needs.

⁴ [WFP Evaluation Policy 2022](#).

On an exceptional basis, the Office of Evaluation or its technical partners can manage a small number of “priority” impact evaluations that fall outside the IEWs if these are of corporate interest to WFP globally. Non-window impact evaluations should be fully funded by the country office or requesting donor, including the costs of support provided by the Office of Evaluation and its technical impact evaluation partners. Any offices considering an impact evaluation outside the thematic windows should consult their regional evaluation officer and the Office of Evaluation’s Impact Evaluation Unit.

Figure 2 (below) provides a high-level overview of the typical impact evaluation process in WFP. Objective 3 of the WFP Impact Evaluation Strategy (2019-2026) aims to “Maximize the responsiveness of impact evaluations to rapidly evolving contexts”. To ensure impact evaluations meet this objective, the timing and length of each phase of the process can be adjusted to harness opportunities, maximize usefulness, and align with evolving contexts.

Figure 2: Impact evaluation process overview



ROLES AND RESPONSIBILITIES

All WFP impact evaluations are managed by the Office of Evaluation and conducted by the Impact Evaluation Unit and its technical partners in close coordination with the regional bureau⁵ and country office involved in implementing the programme. Impact evaluations are delivered through a hybrid partnership model, which combines external expertise and in-house capacity to ensure the impact evaluations align with and are responsive to changes in programme implementation. The following section outlines the roles and responsibilities; a detailed table of roles and responsibilities is also presented in Annex II. Figure 3 also provides a visual representation of the different window-level and country-level stakeholders.

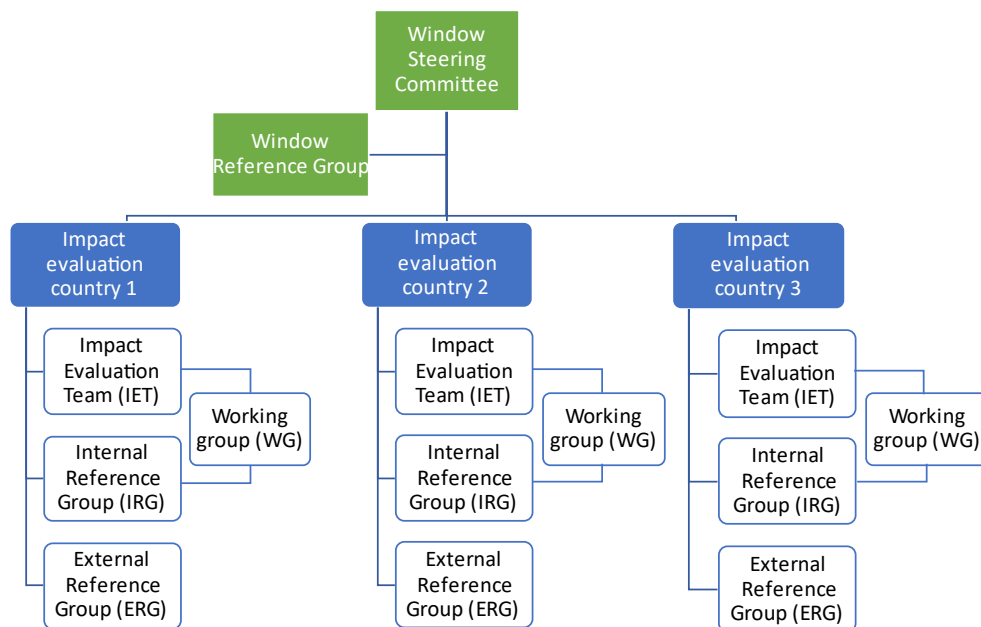
⁵ In their functional role on oversight and technical support to country offices.

Impact evaluations, within an IEW, fit within the wider IEW governance structure. At the window level, the **WSC** is responsible for advising the Office of Evaluation on the formal inclusion of countries in a thematic window. The WSC comprises members from WFP (the Office of Evaluation and other relevant divisions) and research partners (for example, DIME). The WSC typically includes two to five representatives from a relevant WFP division, a minimum of two representatives from the Office of Evaluation (the WC and the Head of the Impact Evaluation Unit) and two from the evaluation partners. The WSC is expected to meet at least once a year, and any time there is a need to discuss the inclusion of an impact evaluation into the window. The WSC is chaired by the Head of the Impact Evaluation Unit, with the WC serving as secretary.

The **WC** is a WFP impact evaluation officer who serves as the interlocutor for the steering committee and the country offices. The WC is responsible for ensuring applications to the window are complete and that relevant documents are shared with the steering committee for country selection.

At window level, the **Window Reference Group** serves as the first point of contact for global engagement efforts and outreach events associated with the IEW products. The Window Reference Group typically includes members from organizations with an interest in the thematic area covered by the window. It may also include academics, technical expertise and inter-agency representation, as relevant. The Window Reference Group is expected to meet at least once a year or more, as relevant.

Figure 3: Window- and country-level governance arrangements*



*The evaluation manager is part of the IET.

At **individual impact evaluation level**, the governance structure includes:

- the IET;
- the Internal Reference Group (IRG);
- the Working Group (WG) (country office programme and research, assessment and monitoring focal points); and
- the External Reference Group (ERG) (optional).

The **IET** is responsible for designing, managing and delivering the evaluation throughout all its steps and maintaining relationships with the country office and governing bodies. The IET includes an EM, technical lead, research analyst and FC. In general, the EM is based at the Office of Evaluation, while the technical lead and research analyst are from an external partner (e.g. DIME). The IET may also include external academic partners (for example, local or regional academic researchers and qualitative researchers).

The **EM** is a WFP impact evaluation officer. S/he is responsible for overall implementation throughout the evaluation process and for ensuring that the evaluation responds to WFP evidence priority needs. The EM provides the first level of quality assurance. In line with the [WFP Evaluation Policy 2022](#), the Office of Evaluation EM can also play a more significant role in an evaluation, such as that of team leader, who is responsible for the overall technical quality of the evaluation. All IET members shall not have a vested interest in the evaluand (i.e. the subject under evaluation). In cases of WFP staff, they should come from an independent evaluation unit with clear and distinct career paths and career progression incentives that are different from the programme's performance.

The **FC** is based at either the country or regional level and is responsible for liaising with the programme team and cooperating partners throughout the impact evaluation. Generally, the FC is based in the country where the intervention is implemented. S/he should allocate 80 to 100 percent of their time to the impact evaluation. In cases where the country office or regional bureau hires the FC to be part of their monitoring and/or evaluation team, the other 20 percent of their time is usually focused on supporting programme-related data collection activities. Arrangements for how the FC is recruited will vary on a country-by-country basis, as they can be hired by WFP's country office or the evaluation partner(s). The FC will need to have access to the field and WFP data and information, including access to WFP systems and the WFP duty of care.

While impact evaluations are centrally managed by the Office of Evaluation, the **country office and cooperating partners** implementing the intervention are the key stakeholders of the evaluation. They are expected to be closely involved in impact evaluation design, ensure correct implementation of the intervention, inform the IET of changes or delays in implementation, and provide co-financing and logistical support for data collection.

The impact evaluation **WG** is the key interlocutor between the IET and the country office during the entire impact evaluation process:

- The focal points for the impact evaluation are agreed at the time of signing the memorandum of understanding (MoU). It is recommended that a member of the country office programme team and the vulnerability analysis and mapping, and monitoring and evaluation teams are members of the WG.
- The WG is responsible for ensuring that the programme and intervention(s) are implemented as outlined in the impact evaluation Concept Note (internal) that is attached to the signed MoU.
- The WG is expected to engage regularly: depending on the phase, this can range from weekly to bi-weekly to monthly.
- The FC (or research analyst) coordinates the WG meetings and ensures effective communications.

The **IRG** is chaired by the country director or their designee (e.g. the deputy country director or Head of Programme), with the EM serving as secretary. It is expected to meet during each evaluation phase. Membership of the IRG is determined by the country office and the Office of Evaluation and can include regional bureau representatives. The IRG reviews key outputs during each phase of the impact evaluation. As regional evaluation officers and regional programme advisers are also important sources of advice and support during the selection, management and delivery of impact evaluations, they can be invited to join the IRG.

The **ERG (optional)** may be established on a case-by-case basis. The IRG will determine whether an ERG is useful and necessary. The purpose of the ERG is to involve internal and external stakeholders and build ownership in the evaluation process and results to maximize evidence use and uptake. External stakeholders might include beneficiaries, governments, and other humanitarian and development actors such as national and international non-governmental organizations, donors and other UN agencies. Members of the ERG are identified by the IRG, which also defines the governance structure and meeting frequency based on needs. While the ERG is not a compulsory governing body, it is highly recommended for donor-funded impact evaluations or when there is a clear policy audience for the evaluation.

Phase 1: Planning

The principles of utility have specific implications at the planning phase as follows:

Utility: An explicit discussion on the intended use of the evaluation takes place with internal and external stakeholders. The evaluation questions should be framed within one of the existing impact evaluation windows/workstreams and contribute to established evidence needs and/or to programmatic operational needs.

Ethics: The *UNEG Ethical Guidelines for Evaluation* include a pledge of ethical conduct in evaluation that has to be followed by all evaluation commissioners and evaluators contracted by WFP, as well as a checklist of ethical issues that the EM and evaluation teams should consider at each phase of the process.

Initiating an impact evaluation. There are two ways that a country office can initiate a request to the Office of Evaluation for a country-level impact evaluation: 1) through an IEW; or 2) by contacting the Office of Evaluation’s Impact Evaluation Unit (copying the relevant Regional Evaluation Unit) to explore interest in conducting an impact evaluation. Office of Evaluation support for any impact evaluation is approved by the Director of Evaluation upon request from and consultation with the country director in the country concerned, when the Director of Evaluation and country director sign an MoU outlining their commitment to the impact evaluation.

The first, and most common, route to initiate a country-level impact evaluation is through an IEW. Each year, the Office of Evaluation circulates a note outlining the thematic focuses and priority questions for each open window to regional bureaux and country offices, together with a request for an **expression of interest**. The Concept Note is intended to inform discussions with key stakeholders at an early stage of the IEW on the thematic scope and potential evaluation questions of interest for the IEW. The expression of interest is the country office’s primary pathway for consideration into the IEW. All country offices that submit an expression of interest are required to participate in an impact evaluation training course, delivered by the Office of Evaluation, aimed to ensure a common understanding about key concepts and processes.⁶

The estimated timing for this phase varies greatly depending on whether the proposed impact evaluation would be part of an already established IEW or a new one. Established IEWs have already reviewed the existing literature and identified relevant questions, associated designs and evaluation tools.

Budgeting and funding for impact evaluations. Impact evaluation costs vary significantly depending on impact evaluation design (e.g. questions, number of intervention arms, sample size, number of data collection rounds, etc.) and the country and programme contexts (e.g. accessibility of sites, security situation, etc.). Table 2 provides an estimate of the cost range for impact evaluations, including technical assistance and data collection costs. On average, an impact evaluation cost is around USD 500,000. The Office of Evaluation’s Impact Evaluation Unit will support in costing estimates based on the questions and designs available.

Table 2: Indicative impact evaluation costs

Cost category	Range	
	Low	High
Technical assistance (includes technical leads, evaluation specialists, analysts, FCs)	USD 100,000	USD 1,200,000
Data collection (includes the survey teams, logistics, equipment, etc.)	USD 200,000	USD 800,000
Indicative totals	USD 300,000	USD 2,000,000

⁶ Relevant country office and regional bureau staff (heads of programmes or the unit leading implementation, programme officers, monitoring and evaluation staff, and staff in sub-offices) are all invited to join the training.

An IEW can accommodate up to six ongoing country-level impact evaluations. The Office of Evaluation and its partners engage in discussions with country offices to determine which country-level impact evaluations present the highest chances for inclusion into the IEW. If an impact evaluation is accepted within an IEW, it will receive co-financing from the Office of Evaluation.

The country office should reflect the full cost (i.e. Office of Evaluation and country office funding) of the impact evaluation in the country portfolio budget and the vulnerability analysis and mapping, and monitoring and evaluation evidence planning and budgeting tool. The country office/regional bureau should also allocate 10 percent of support time to member(s) of the WG, 15 percent to the focal point, and 5 percent to member(s) of the Evaluation Management Group.

The country office will work jointly with the Office of Evaluation's Impact Evaluation Unit, regional evaluation officers and research partners to identify funding sources and secure adequate funds. This may be during the country strategic plan visioning workshop discussions, taking place every four to five years in each country; during the annual updates of the evidence planning and budgeting exercise or regional evaluation plan (REP) consolidations; or finally, it may take place during a specific programme funding proposal phase.

Assessing feasibility of an impact evaluation. Feasibility of an impact evaluation is determined through a feasibility assessment process. The [Feasibility Assessment Form](#) helps guide conversations on whether a proposed impact evaluation is feasible. This form is aimed at guiding the discussion and documenting whether it is feasible to design a rigorous impact evaluation for an intervention. Typically, feasibility assessment is conducted at least three to six months prior to the activities' launch. The form is expected to be completed by the Office of Evaluation's Impact Evaluation Unit together with the country office and regional evaluation officer. Considerations to deem an intervention feasible for a rigorous impact evaluation include programme/pilot characteristics, resource and capacity considerations, and demand for and use of evidence. Not all interventions are suitable⁷ for an impact evaluation, in which case the WSC will not approve the country office's inclusion into the IEW, and the impact evaluation process ends. Also, in most cases, other evaluation methods are feasible and can be explored through a decentralized evaluation when an impact evaluation is not possible, and related discussions can continue with the regional evaluation officer.

⁷ Impact evaluations may not be feasible in cases where the number of units (e.g. communities, schools, households or individuals) is small, as this makes it impossible to detect any statistically significant differences. Other challenges can include the timing of the evaluation (e.g. whether there is enough time to prepare for data collection before an intervention), the timing of expected outcomes (e.g. whether the outcomes will manifest during the programme period) and budget constraints (e.g. insufficient data budgets), among others.

Box 1: Technical feasibility

Technical requirements. As part of the planning for an impact evaluation, the following technical requirements must be met:

- A well-defined and testable **theory** is essential for identifying how the programme or project is intending to achieve its causal impacts. This will help the process of agreeing on appropriate evaluation questions. The Office of Evaluation and its technical partners will support the refinement of a testable theory and/or theory of change.
- **Scale and timeframe** of programme implementation is large enough (scale is typically no smaller than 30 clusters, depending on what the units of intervention are) and long enough to be able to detect the intended impact on one or more target outcomes. Timeframe depends on the nature of the intervention and outcomes.
- **Well-specified counterfactual** (i.e. what would have happened in the absence of the intervention). This should be a population group that was not exposed to the intervention (or was exposed to a different type of intervention) but that was otherwise identical and on the same development trajectory as the intervention group. Under specific circumstances (which are not often met) a counterfactual can be constructed using statistical methods and can still provide a credible impact evaluation.
- **Reliable quantitative and qualitative data** collected using the right sample size (based on power calculations), at the right level (e.g. household or individual) and using the appropriate timeframe (e.g. before, during and after interventions). The Office of Evaluation and its technical partners will lead survey design, data collection, and data analysis processes.

Utility: During the feasibility assessment, the country office and the Office of Evaluation will also examine the usefulness of any potential impact evaluation. Impact evaluations are used for learning in WFP, and it should be clear that the evidence generated is novel (e.g. existing literature is not available) and can inform future decisions (e.g. there is an expectation that the type of intervention will be used again). The usefulness of the evidence generated will also inform discussions with the WSC and influence its recommendation for the impact evaluation to be included in any window.

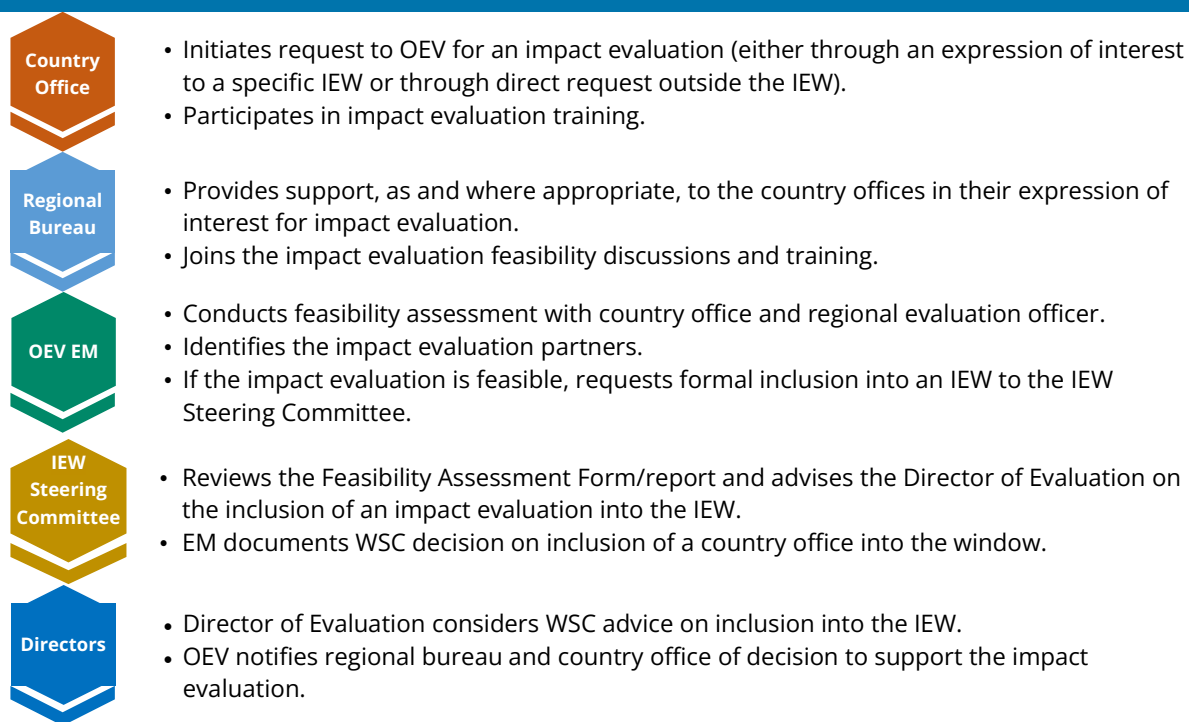
Ethics: In addition to technical feasibility and utility, careful consideration will be made regarding the ethics of conducting any impact evaluations, including the ethics of the design (e.g. comparison groups), the process (e.g. how people will be targeted or included in the impact evaluation), and the data collection and management. Following the feasibility assessment, all impact evaluation designs undergo an independent ethical review by an accredited IRB during the preparation phase (see “Phase 2: Preparation” below).

The WSC will review the feasibility assessments of all the proposed country-level impact evaluations and endorse inclusion into the IEW. The Director of Evaluation will then approve, and the Office of Evaluation will write to the country director to officially inform them about the inclusion into the IEW. Windows can remain open as long as the topics covered are a priority for WFP. New countries may be added to an IEW only after one of the country-level impact evaluations is completed.

1.1 PLANNING PHASE PROCESS GUIDE

Identifying the partners and establishing the IET. Impact evaluations are delivered through a hybrid partnership model, which combines external expertise and in-house capacity. During this phase, the Office of Evaluation will explore and identify the partner and required expertise to conduct the impact evaluation and establish the IET.

Figure 4: Process map for the planning phase



1.2 PLANNING PHASE OUTPUTS

The outputs for the planning phase are:

- ✓ feasibility assessment;
- ✓ endorsement from the WSC on inclusion of impact evaluation into the window; and
- ✓ Director of Evaluation communication informing country offices of inclusion into the window and next steps.

1.3 PLANNING PHASE CONTENT GUIDE AND QUALITY STANDARDS

The purpose of this section of the guidance is to assist the IET in completing the feasibility assessment for a country office's inclusion into an IEW. The [Technical Note on Quality Standards for Impact Evaluations](#) provides detailed guidance to assess whether an impact evaluation is feasible and obtain approval to officially join an IEW. The Feasibility Assessment Form serves as a checklist to determine whether the right conditions are in place to initiate an impact evaluation. There is no word length requirement.

1.4 REFERENCE MATERIALS FOR PLANNING PHASE

Templates and quality checklists	Technical notes	Other reference material
<ul style="list-style-type: none"> ✓ Feasibility Assessment Form 	<ul style="list-style-type: none"> ✓ Quality Standards for Impact Evaluations 	<ul style="list-style-type: none"> ✓ Impact Evaluation Decision Guide ✓ OEV Communication Protocol

Phase 2: Preparation

The principles of independence, impartiality, credibility and utility, as well as ethics, have a number of specific implications in the preparation phase:

Independence: The EM ensures that the evaluation team members have no vested interest in the subject to be evaluated.

Impartiality: The EM and technical lead prepare all the required documents (MoU, risk matrix, ethical clearance) following this process guide to ensure absence of bias in terms of scope and design. The Programme Evaluation Committee, which includes key stakeholders, is formed to help steer the evaluation and reduce risk of bias. Impartiality is critical to the usefulness and credibility of the evaluation.

The **credibility** of an evaluation is determined by its methodological appropriateness and rigour. The IET must have a track record of producing high-quality impact evaluations under the subject of analysis. Independence, impartiality and transparency are also key components of a credible evaluation.

Utility: An explicit discussion on the intended use of the evaluation takes place with internal and external stakeholders at country level. The country-level evaluation questions should be focused, appropriate and relevant to the users' needs and linked to the evaluation's objective(s). The EM effectively steers the evaluation process to ensure it meets WFP evidence needs.

Ethics: The evaluation design needs to be approved by a recognized IRB. All IET members who handle personally identifiable information must have up-to-date Human Subjects Research Certification and every three years must complete the Protecting Human Research Participants training. Evaluation design should be such that participants are not denied an intervention, treatment or services they would otherwise be entitled to receive.

The preparation phase begins after completion of the planning phase, when Office of Evaluation support for an impact evaluation is approved in response to a request for support from the country director in the country concerned, and in most cases the inclusion into an IEW. As impact evaluations are managed through a hybrid model, the governance structure is broadly outlined in [Roles and Responsibilities](#) and is detailed in an MoU between the Office of Evaluation, the evaluation partner(s) and the country office.

The EM is responsible for setting up the IET and agreeing with the country office the composition and establishment of the IRG. Once established, the IRG decides whether an ERG is useful and necessary. The EM supports the recruitment of an in-country [FC](#), or field manager, who is usually based in the WFP country office.

The EM registers the impact evaluation in the **Office of Evaluation Management Information System** and sets up a dedicated Teams space to house the **Evaluation Library** as the repository for documentation relevant to the evaluation that is to be shared between the IET and the IRG.

The IET conducts a **design workshop mission** to the country office with the programme team, IRG and other relevant key stakeholders' participation. The design workshop focuses on better understanding of the intervention context, constraints and features in order to inform the impact evaluation Concept Note. The design workshop also serves to determine if a pilot is required (outlined below).

The IET develops a [Concept Note](#), an internal document that outlines the purpose, scope, evaluation design, methodology and sampling strategy, sequencing and frequency of quantitative and qualitative data collection, timelines, roles and responsibilities. The Concept Note (internal) is then attached to the MoU as an appendix and informs the Inception Note (external).

The IET, together with the WG, completes a [risk matrix](#), highlighting potential risks and associated mitigation strategies. The risk matrix is attached to the MoU as an appendix and informs the Concept Note.

The IET finalizes the evaluation design, including the data collection tools and **pre-analysis plan** (PaP). The IET registers the design in a PaP (if not part of the window's PaP) on an internationally recognized evaluation/research registry (e.g. [AEA RCT Registry](#), [RIDIE](#), etc.). The PaP is intended for an external

academic audience, and it describes how the IET plans to analyse the data in advance of data collection, including the regression models.

In order to meet International Compilation of Human Research Standards, the IET applies for clearance from an international **IRB** and relevant national IRBs to ensure the protection of human subjects as part of the impact evaluation.

A key deliverable of the preparation phase is an **MoU**, which is developed by the Office of Evaluation, the evaluation partner and the country office. The MoU includes the roles and responsibilities throughout the evaluation; the activities that will take place under the evaluation; financial arrangements to cover the impact evaluation costs; the data-sharing agreement for the impact evaluation; the agreement to hire an FC placed in the country to coordinate evaluation activities; and the country office's plans to use the evidence. The Concept Note (internal) and risk matrix are attached to the MoU as appendices. The MoU is approved and signed by the Director of Evaluation, country director and evaluation partners, and then enables the impact evaluation to move into the inception phase.

2.1 PREPARATION PHASE PROCESS GUIDE

The purpose of this section of the process guide is to provide a step-by-step description of the process leading to the finalization of the Concept Note, risk matrix and MoU; and the establishment of the evaluation team, Programme Evaluation Committee and ERG (optional), highlighting the roles and responsibilities of each stakeholder.

The steps, including the roles, responsibilities and actions, are provided in Figure 5.

Figure 5: Process map for the preparation phase



2.2 PREPARATION PHASE OUTPUTS

The outputs for the preparation phase are:

- ✓ Office of Evaluation Management Information System registry;
- ✓ terms of reference for FC or field manager;
- ✓ Teams space (Evaluation Library);
- ✓ signed MoU;
- ✓ risk matrix;
- ✓ Concept Note (internal);
- ✓ IRB approval; and
- ✓ design/PaP registered (if not part of the window's PaP).

2.3 PREPARATION PHASE CONTENTS GUIDE AND QUALITY STANDARDS

The purpose of this section is to assist the EM in drafting the Concept Note (internal) and to guide stakeholders in understanding the contents of the workplan for an impact evaluation. The Concept Note (internal) should follow the structure and standard content described in the [Concept Note template](#). The Concept Note is developed by the IET and reviewed by the IRG and the Head of the Impact Evaluation Unit

prior to being attached to the MoU that is reviewed and signed by the Director of Evaluation and country director.

While the contents will be adapted to the context and specific subject under evaluation, the Concept Note must contain core contents:

- **context** within which the impact evaluation is to be conducted, including geography, demography, food and nutrition security, and appropriate internal and external institutional context;
- **programme description** of the intervention being evaluated, including scale and scope, main elements and hypothesized causal chain (theory of change if it exists);
- **evaluation questions** connecting the intervention with the outcomes of interest and how these were delivered (may include additional/secondary questions on implementation);
- **evaluation design**, including identification strategy (comparison groups, intervention arms, inclusion/exclusion criteria, etc.) for each evaluation component, and statistical methods that will be used to compare groups for all outcomes included in the evaluation;
- **data and measurements**, including data sources, collection instruments, sampling strategy, sample size and outcome indicators;
- **ethical considerations**, including IRB approval, informed consent procedures, confidentiality, and transparency in the design (including experimental considerations, gender considerations, etc.);
- **evaluation governance**, including evaluation team, management group, WG, reference groups, and their functions/roles/responsibilities within the specific evaluation context; and
- **evaluation workplan** with the phases and estimated timelines.

2.4 REFERENCE MATERIALS FOR PREPARATION PHASE

Templates and quality checklists	Technical notes	Other reference material
<ul style="list-style-type: none"> ✓ Risk matrix template ✓ Concept Note template ✓ MoU template 	<ul style="list-style-type: none"> ✓ Technical Note on Principles, Norms and Standards ✓ Technical Note on Evaluation Questions and Criteria ✓ Technical Note for Planning and Conducting Evaluations During COVID-19 ✓ Technical Note on Stakeholder Analysis ✓ Technical Note on Integrating Gender, Quick Guide and Checklist 	<ul style="list-style-type: none"> ✓ FC terms of reference ✓ Example Impact evaluation MoU

2.5 IDENTIFICATION OF THE COUNTERFACTUAL

During the counterfactual phase, the principles of independence, credibility and ethics are applied.

Independence: Identification of the counterfactual and randomization is conducted by the evaluation team, or under its strict supervision.

Credibility: The randomization process should be open and transparent.

Ethics: The evaluation team establishes communication protocols for engaging with local communities, presenting the scope and process of the programme and evaluation.

For an impact evaluation design to be successful, there needs to be a credible **counterfactual**. A counterfactual is used to measure what would have happened in the absence of the intervention. Since counterfactual outcomes are unobservable, they need to be estimated from the observed outcomes of the comparison group. Counterfactuals can be constructed using different experimental and quasi-experimental evaluation designs. An experimental design (or randomized controlled trial) develops a counterfactual through the process of randomization.

In cases where randomization is not possible or feasible, there are several quasi-experimental designs to develop a counterfactual using a comparison group that has not been created by randomization.

Described below and in the [Technical Note on Quality Standards for Impact Evaluations](#) are the conditions, tests and assumptions that need to be met for quasi-experimental designs to provide a credible counterfactual.

- Regression discontinuity design minimum requirements:
 - Programme is allocated on a needs basis or means-tested basis, and limited programme budget implies that many individuals/households/schools/villages above the threshold cannot be served.
 - There is no selective targeting manipulation.
 - Design passes validity and density tests.
- Difference-in-difference design minimum requirements:
 - Data are available for at least two pre-intervention periods before the intervention begins, to allow for test of the parallel trend or test of the common trend assumption.⁸
- Propensity score matching design minimum requirements:
 - Data are available for at least two pre-intervention periods before the intervention begins, to allow for placebo treatment tests.
 - Absence of programme selection based on unobservable variables.
- Instrumental variables design minimum requirements:
 - Existence of an instrument that is correlated with programme participation but uncorrelated with the outcomes.
 - Over-identification test and a credible argument for the absence of a direct effect.

2.6 PROCESS GUIDE FOR IDENTIFICATION OF COUNTERFACTUAL PHASE

This section outlines the process for assigning intervention and comparison groups through randomization.

The WG will identify and share with the IET the list of all potentially eligible units (e.g. areas, communities, households, individuals, etc.) where the intervention might be implemented. This is typically done based on

⁸ Matching designs and difference-in-difference designs with one data point as baseline for both intervention and comparison group may be considered a medium-quality design approach.

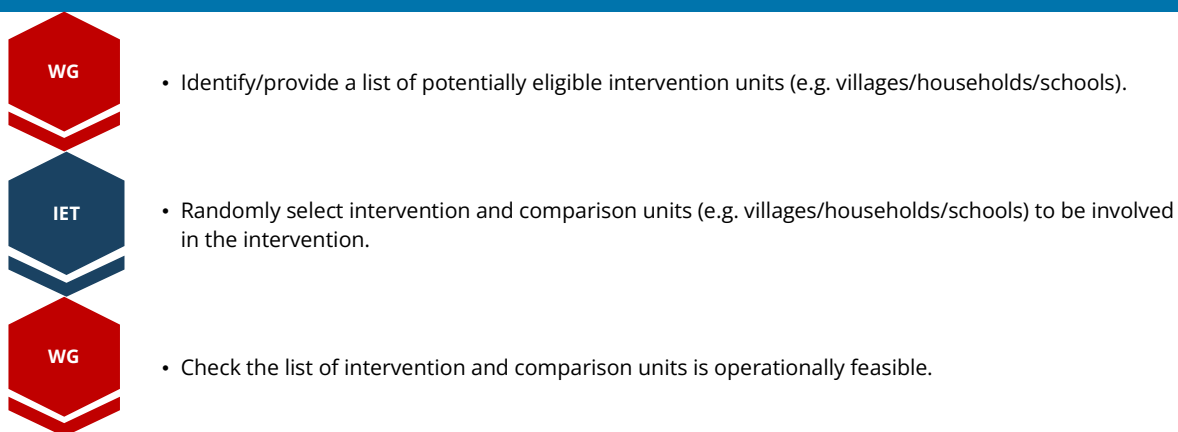
a well-defined targeting strategy and set of characteristics to determine eligibility (for example, all vulnerable households in defined provinces). The IET can support in this process if necessary.

Once the IET has received the list from the WG, it will conduct the randomization process. During randomization, units are randomly assigned to one or more groups that receive one of more types of an intervention. In its simplest form, this process will create an **intervention group**, which receives the intervention, and a **comparison group** that does not receive the same intervention (or receives a different type of intervention). WFP impact evaluations never prevent people from receiving support, and therefore cannot guarantee a pure control group in the traditional sense.

The IET will share the list of intervention and comparison units with the WG, which will review to ensure that it is operationally feasible. The WG must ensure that intervention and comparison groups are maintained through the intervention as outlined in the Concept Note, or in the case of a pilot, the pilot design report.

Randomization can take place any time between the preparation and before the implementation phase.

Figure 6: Process map for the randomization phase



2.7 IDENTIFICATION PHASE OUTPUTS

The outputs for the randomization phase are:

- ✓ list of potentially eligible intervention units; and
- ✓ list of sites divided by intervention and comparison group.

2.8 REFERENCE MATERIALS FOR RANDOMIZATION PHASE

Templates and quality checklists	Technical notes	Other reference material
	Quality Standards for Impact Evaluations	

2.9 OPTIONAL PILOT IMPACT EVALUATION (LEAN APPROACH)

The principles of independence, credibility and utility, as well as ethics, have a number of specific implications for pilot impact evaluations as follows:

Independence: Ensuring independence and impartiality means the non-participation by WFP programme or partner organization staff in the IET's data collection. The IRG ensures that the IET has full access to available information and data as per the WFP directive on information disclosure. Evaluators have full freedom to conduct their evaluative work without interference or fear for their career. Any challenges should be reported to the Director of Evaluation in a timely manner to facilitate resolution.

Credibility: Data should be collected in both intervention and comparison groups using the same tools. Importantly, questionnaires should refrain from directly asking about participation in WFP programmes and should be structured to avoid courtesy bias. If challenges arise during the field mission, adjustments are made, ensuring that those do not undermine impartiality.

Utility: The EM ensures that the pilot is implemented to serve the evaluation evidence needs. The EM organizes a discussion on the preliminary findings at the end of the pilot analysis and the participation of relevant country office/regional bureau/headquarters staff.

Ethics: The evaluators behave ethically in all interactions with stakeholders and beneficiaries. The evaluation team establish clear protocols and procedures for dealing with data collection of sensitive and personal data that are in line with the most updated rules and regulations. They ensure informed consent, protecting privacy, confidentiality and anonymity of participants and cultural sensitivity. They respect the autonomy of participants. They ensure the participants' fair recruitment, and that pilot survey results do not harm them or their communities. They have an obligation to report any noted incidents of fraud, corruption, Protection from Sexual Exploitation and Abuse (PSEA) or other wrongdoing.

During the preparation phase, the Office of Evaluation and the country office will determine whether a pilot impact evaluation is required. The Office of Evaluation recommends using a pilot impact evaluation in cases where the time and scale of programming are limited and aspects of the programme are yet untested, and/or the impact evaluation approach needs to quickly respond to timing and context (e.g. humanitarian responses, etc.). Office of Evaluation support for pilot impact evaluations is approved by the Director of Evaluation upon request from the country director in the country concerned. Pilot impact evaluations are selected for Office of Evaluation support based on the potential value of evidence generated for WFP programmes, and/or the potential for future full-scale impact evaluations to be included in an IEW.

If during the feasibility assessment it emerges that aspects of the intervention need to be tested, a small-scale pilot will be conducted during the preparation phase to provide confidence in the **feasibility** of a future impact evaluation design for evaluating an intervention. While a pilot is optional, it is highly recommended for new interventions with limited implementation experience and/or timescale. The objective of the pilot phase is to test whether there is enough implementation capacity prior to the launch of a large-scale impact evaluation.

A pilot often includes a smaller-scale data collection process on the final outcome to test data collection tools prior to the launch of the baseline and provide more precise estimates for power calculation analysis.

Typically, a pilot uses a [lean impact evaluation](#) approach, which allows the IET to compare two (or more) options for addressing operational challenges. Lean impact evaluations can be conducted using an experimental design with unit random assignment, to test alternative implementation modalities without needing a comparison group. This has the advantage of familiarizing the programme team with the randomization process on a smaller scale before full implementation. However, rather than focusing on outcomes, lean impact evaluations focus on comparing output level data and mainly rely on already existing monitoring systems for data collection. This has the advantage of minimizing data collection costs, while providing reliable evidence on the implementation of alternative intervention models. Data on final outcomes collected during a pilot are not large enough to make any causal claim.

The pilot will conclude with a Pilot Brief that outlines the findings, including an updated feasibility assessment advising on whether or not to proceed with the impact evaluation.

2.10 PILOT IMPACT EVALUATION PROCESS GUIDE

A pilot impact evaluation proceeds through eight steps:

- 1 **Pilot design:** The IET, in close collaboration with the IRG and cooperating partners, develops a [pilot Concept Note](#), which will be reviewed and cleared by the IRG. The pilot Concept Note is an internal document.
- 2 **Pilot randomization:** The WG shares the list of all potentially eligible units (e.g. villages/households/schools) to be involved in the pilot. The IET randomizes the sample from this list, identifying the intervention and comparison groups. The sample list is then shared back with the programme team to begin the pilot implementation with the selected units. Phase 3 provides a more detailed outline of the randomization process.
- 3 **Pilot implementation:** The programme team implements the pilot as outlined in the pilot design report. Implementation typically lasts three to nine months.
- 4 **Pilot data collection:** The IET works with the WG to hire a data collection firm or establish alternative arrangements to set up a data collection team. The FC liaises with the IRG to set up a data collection process for monitoring and survey data. The IET conducts data collection training with enumerators and/or cooperating partners (either recruited directly by WFP or through an external data collection company identified by the country office or evaluation partners) and sets up the quality check system to monitor data quality.
- 5 **Pilot data analysis:** The IET cleans and conducts data analysis on pilot data.
- 6 **Discussion of pilot findings:** The IET presents and discusses with the IRG the findings from the pilot analysis.
- 7 **Pilot Brief:** The IET produces a Pilot Brief in cases where a full impact evaluation is not feasible or if requested by the country office. The note will include the design, findings and recommendations on whether to continue with the impact evaluation. The Pilot Brief will undergo the quality assurance process described below in the content guide and quality standards.
- 8 **Scale-up workshop:** In the event that the Pilot Brief determines that an impact evaluation is feasible, the IET will conduct a scale-up design workshop together with the programme team, IRG and other relevant key stakeholders. The workshop will give the opportunity to inform changes in the evaluation and programme design, based on the evidence and learning coming from the pilot.

The outputs for the pilot impact evaluation (lean approach) are:

- ✓ pilot Concept Note;
- ✓ Pilot Brief; and
- ✓ revised feasibility assessment (in cases where intervention scale-up is planned).

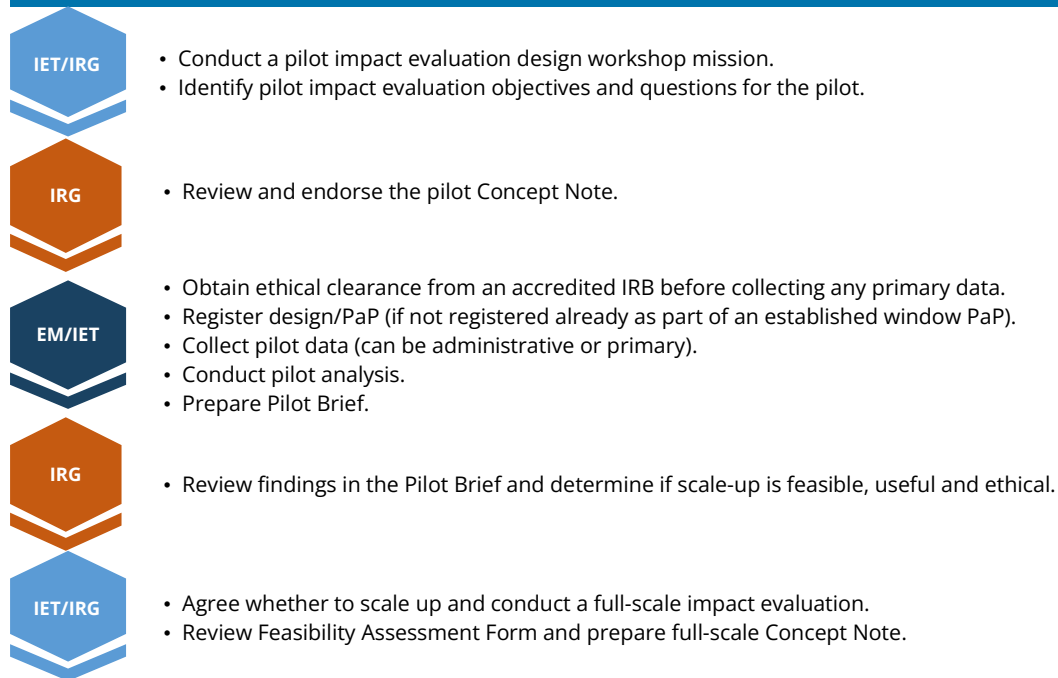
2.11 CONTENT GUIDE AND QUALITY STANDARDS FOR PILOT IMPACT EVALUATIONS

The purpose of this section of the guidance is to assist the EM in drafting the Pilot Brief. The Pilot Brief should follow the structure and standard content described in the template.

Data should be presented in a clear and concise manner (in tables, diagrams, etc.) as appropriate for effective communication. They should be systematically analysed and interpreted. Findings should be evidence-based and relevant to the evaluation questions under review. The IET should make a clear distinction between facts borne out by evidence and assumptions or plausible associations they draw from the evidence. Conclusions should follow logically from the analysis of data and findings. The Pilot Brief should be balanced and impartial and use constructive language.

The EM, with the support of the evaluation analyst, carries out first-level quality assurance of the pilot note, using the quality checklist. The EM shares the revised Pilot Brief with the IRG and ERG, if applicable. The revised Pilot Brief is then reviewed by QA1, QA2 and the Head of the Impact Evaluation Unit (if different from QA2).

Figure 7: Process map for the pilot phase



2.12 REFERENCE MATERIALS FOR PILOT IMPACT EVALUATION

Templates and quality checklists	Technical notes	Other reference material
<ul style="list-style-type: none"> ✓ Pilot Concept Note template ✓ Review Matrix 	<ul style="list-style-type: none"> ✓ Technical Note on Principles, Norms and Standards 	-

Phase 3. Inception

The principles of independence, impartiality, credibility and utility, as well as ethics, have a few specific implications at the inception phase as follows:

Independence: To ensure the adequate implementation of evaluation standards and principles, the Evaluation Management Group should provide the evaluation team with access to all available data and programme information.

Credibility: The evaluation methodology should be appropriate to respond to the evaluation questions, and the methods for data collection, analysis and interpretation should be transparently documented in the Inception Note. The Inception Note should benefit from inputs from key stakeholders and a rigorous quality assurance process, including inputs from external anonymous peer reviewers.

Utility: The IET should ensure an efficient evaluation process as per the timeline in order to avoid late completion of the evaluation. Changes to the timeline to fit with programme implementation requirements are discussed and agreed with the IRG and IET.

Ethics: Evaluators should behave ethically in all interactions with stakeholders. Adequate ethical safeguards and considerations should be indicated in the Inception Note.

The inception phase begins when the risk matrix, Concept Note and MoU are completed. If a pilot is conducted, inception will take place when the pilot is completed.

The inception phase serves to ensure that the IET develops an in-depth understanding of the subject of the evaluation and can document this in an **Inception Note (optional, external)**. Using the Concept Note as a foundation, the Inception Note describes the proposed evaluation approach and methodology, including the data collection and analysis methods selected to answer the evaluation questions, stakeholder analysis and process steps.

The inception phase builds on the design workshop (and pilot where applicable) to give the IET and IRG a greater understanding of issues and concerns related to the baseline study and subject being evaluated and programme implementation. The Inception Note should further expand elements of the Concept Note, notably ensuring that the evaluation subject, context and scope are correct, relevant, up-to-date, appropriately nuanced and politically sensitive. The Inception Note is meant to clearly confirm the scope, how the work is to be performed, who is to do what, what is to be produced and when deliverables are expected, as well as how deliverables will be disseminated and used. The Inception Note is the main output for this phase.

Within this framework, the main objectives of the inception phase are to:

- refine the context analysis within and external to WFP;
- situate the impact evaluation within the IEW and in relation to other impact evaluations;
- deepen and finalize the evaluation approach, hypotheses, evaluation questions and outcomes of interest;
- fine-tune the evaluation design and sampling strategy as relevant and appropriate, and also in view of pilot findings if applicable;
- define the evaluation methodology, management of data quality and implementation monitoring system, giving due consideration to ethical issues, risks and mitigation strategies;
- develop data collection tools and test them as appropriate and feasible;
- building upon all the above, define the analytical framework, outlining data processing, statistical analysis methods and the analysis strategy for qualitative data;
- deepen and finalize the stakeholder mapping and analysis;
- finalize the communication plan for dissemination and use;

- develop a detailed Inception Note with roles and responsibilities for the team and deadlines for each deliverable; and
- outline quality assurance and peer review processes.

Section 3.1 explains the activities that should be conducted during the inception phase; section 3.2 provides guidance on the expected content and quality standards of the Concept Note; and section 3.3 includes links to all relevant reference documents for this phase.

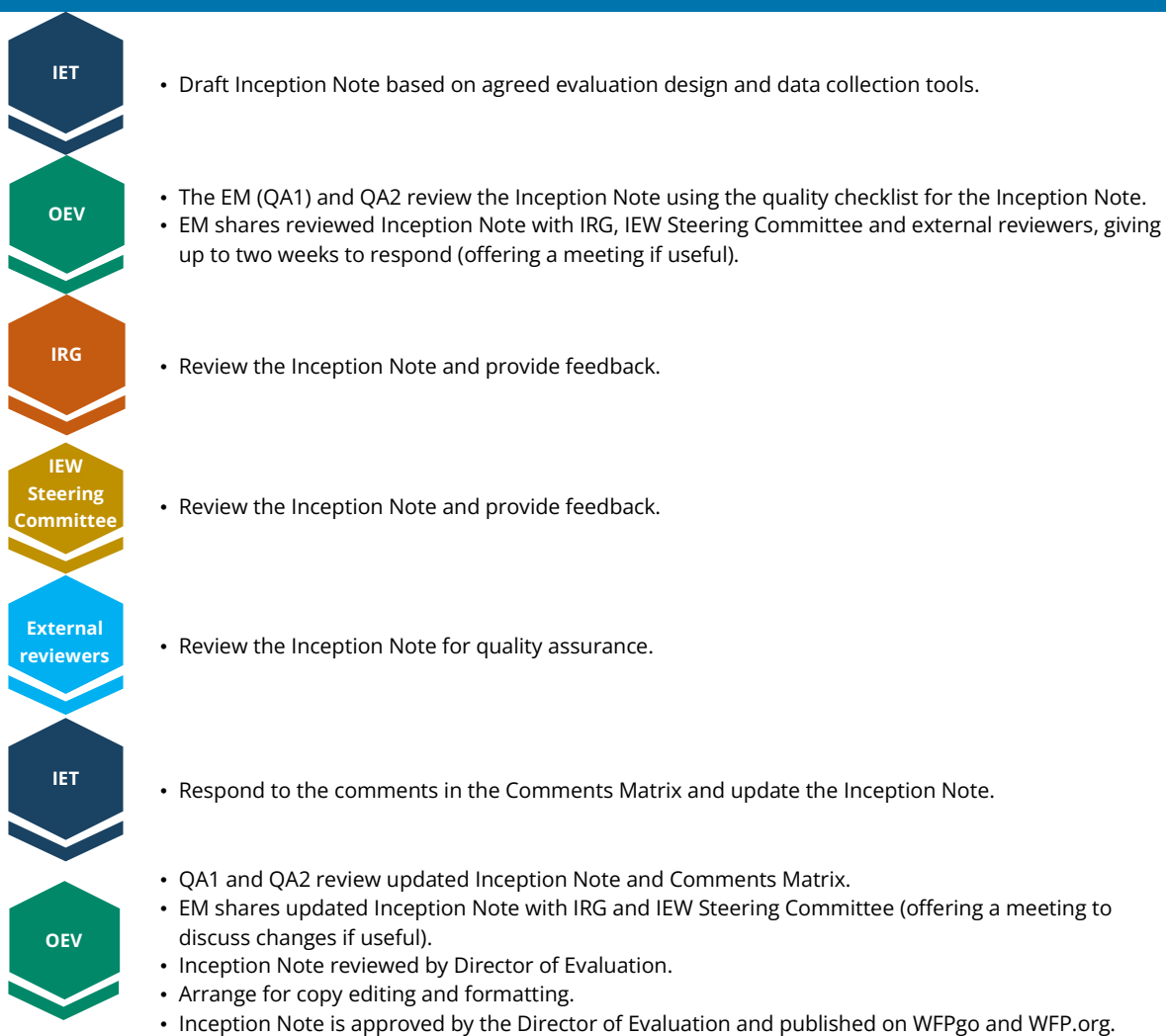
3.1 INCEPTION PHASE PROCESS GUIDE

The process guide clarifies roles, responsibilities and participation during the inception phase and provides a step-by-step description of tasks, particularly those leading to the finalization of the Inception Note for the evaluation.

Preparation of the Inception Note (Optional). The Inception Note provides transparency to key stakeholders on what the evaluation will cover and how it will be carried out. Therefore, the draft Inception Note should focus on issues that affect the evaluation scope, validation of the theory (or theory of change), methodology, fieldwork and analysis. Approaches to data protection, storage and intervention assignment should be clearly identified. All issues should be resolved before the Inception Note is considered final. Ideally, the data collection phase should start after the Inception Note has been approved.

An overview of the roles and responsibilities for different aspects of the inception phase are described in Figure 8.

Figure 8: Process map for the inception phase



3.2 INCEPTION PHASE OUTPUTS

The main outputs of this phase are:

- ✓ Inception Note (Optional); and
- ✓ Comments Matrix with stakeholder comments and how they have been treated.

3.3 INCEPTION PHASE CONTENTS GUIDE AND QUALITY STANDARDS

The purpose of this guidance material is to assist the IET in drafting the Inception Note. The Inception Note should follow the structure described in the [Inception Note template](#). The content should cover the minimum requirements as per the Inception Note quality checklist. It should not exceed 10,000 words, excluding annexes.

Quality assurance aims to ensure that enough research, stakeholder consultations and analysis have been undertaken to confirm the evaluation design and methodology, and to guide its conduct. The impact evaluation Concept Note quality checklist includes:

- criteria concerning the content, especially related to the scope, methodological approach, evaluation matrix, data collection methods and sampling criteria;
- criteria concerning the Inception Note, its feasibility and its likelihood of generating a credible evaluation;
- a check on whether the required content has been included in the Concept Note; and
- process (for example, a timeline).

The EM, with the support of the evaluation analyst, carries out first-level quality assurance of the Inception Note, using the quality checklist to provide systematic and constructive feedback. The EM submits the Inception Note for second-level quality assurance and clearance to the Head of the Impact Evaluation Unit (if different from QA2) to share with the IRG for comments using the [Comments Matrix](#). The revised Inception Note is then reviewed by QA1, QA2 and the Head of the Impact Evaluation Unit (if different from QA2) and sent to the Director of Evaluation for approval.

3.4 REFERENCE MATERIALS FOR INCEPTION PHASE

Templates and quality checklists	Technical notes	Other reference material
<ul style="list-style-type: none"> ✓ Example Inception Note ✓ Comments Matrix 	<ul style="list-style-type: none"> ✓ Technical Note on Principles, Norms and Standards ✓ Technical Note for Planning and Conducting Evaluations During COVID-19 ✓ Technical Note on Stakeholder Analysis ✓ Technical Note on Integrating Gender, Quick Guide and Checklist ✓ Technical Note on Evaluation Matrix 	<ul style="list-style-type: none"> ✓ OEV Communication Protocol

3.5 BASELINE ANALYSIS (OPTIONAL)

The principles of independence, credibility and utility, as well as ethics, have a number of specific implications at the baseline study phase, as follows:

Independence: Ensuring independence and impartiality means the non-participation by WFP programme or partner organization staff in the IET's data collection activities with beneficiaries and external stakeholders. The IRG ensures that the IET has full access to available information and data as per the WFP directive on information disclosure. Evaluators have full freedom to conduct their evaluative work without interference or fear for their career. Any challenges should be reported to the Director of Evaluation in a timely manner to facilitate resolution.

Credibility: The FC ensures that the baseline survey is implemented as per design. Data should be collected in both intervention and comparison groups using the same tools. Importantly, questionnaires should refrain from directly asking about participation in WFP programmes and should be structured to avoid courtesy bias. If challenges arise during the field mission, adjustments are made, ensuring that those do not undermine impartiality.

Utility: The EM organizes a discussion on the preliminary findings at the end of the baseline analysis and the participation of the country office/regional bureau/headquarters.

Ethics: The evaluators behave ethically in all interactions with stakeholders and beneficiaries. The evaluation team establish clear protocols and procedures for dealing with data collection of sensitive and personal data that are in line with the most updated rules and regulations. They ensure informed consent, protecting privacy, confidentiality and anonymity of participants and cultural sensitivity. They respect the autonomy of participants. They ensure the participants' fair recruitment, and that baseline survey results do not harm them or their communities. They have an obligation to report any noted incidents of fraud, corruption, PSEA or other wrongdoing.⁹

Depending on the design of the impact evaluation, a baseline survey is not always required or feasible. Typically, impact evaluations with a short timeframe (i.e. less than two years), or when pilots are conducted, do not require baseline surveys. Instead, the quality of randomization processes is checked using administrative data (e.g. registration data) or monitoring data to check whether the comparison groups are balanced on key observables (e.g. household composition, etc.).

However, in many cases baseline data are collected during the inception phase to ensure that the impact evaluation design is feasible, and to provide the first measurements for comparing changes over time. An analysis of baseline data should always be conducted whenever a baseline survey is carried out prior to the onset of operation activities to establish the pre-operation exposure conditions of the outcome-level indicators. This is the first set of data collected from the intervention and comparison groups and provides an early indication of whether the chosen impact evaluation design is valid in practice, while also gathering useful information about beneficiary characteristics that can inform the programme.

During this phase, the IET collects, synthesizes and starts analysing baseline survey data as indicated in the Concept Note. The details of the baseline study phase are determined by the evaluation design chosen for a given impact evaluation. Therefore, it may differ for each evaluation. The principles provided here apply to all impact evaluations.

If the baseline analysis is completed during the inception phase, the baseline balance tables can be included as an annex of the Inception Note.

3.6 BASELINE ANALYSIS PROCESS GUIDE

In this phase, the operational plan found in the Concept Note is implemented.

The baseline analysis phase is conducted by the evaluation team and consists, in general, of the following steps:

⁹ See guidance here: [Where to seek support and report wrongdoing in WFP | WFPgo](#).

- **Set up data collection team.** The IET works with the WG to hire a data collection firm or establish alternative arrangements to set up a data collection team. The FC will organize enumerator training and support the organization of field site visits and end of fieldwork debriefing to ensure that they are as effective as possible.
- **Sampling.** Using the list of all potentially eligible units (e.g. villages/households/schools) shared by the WG during Phase 3, the IET will perform sampling on the units to be surveyed. The sampling list will then be shared with the data collection team.
- **Baseline data collection.** The IET manages baseline data collection, ensuring data quality and adherence to protocols and standards.

In particular, the FC coordinates in-country activities, including:

- coordinating with the data collection team;
- monitoring the data collection process in line with the requirements set out in the Concept Note;
- preparing for the field site visits;
- providing administrative/logistical support to the data collection team members, including liaising with units/authorities for payments, transport and authorizations as relevant;
- conducting adequate enumerators' training;
- performing piloting of the tools and data collection processes;
- ensuring correct translation of the tools; and
- conducting validity testing to ensure that data collected is of good quality.

Communication and coordination mechanisms between the FC and data collection team should be established for the duration of the data collection phase and made clear to all at the outset.

Data protection and confidentiality. The IET must ensure confidentiality and the protection of data when processing, storing and/or transferring personal data throughout data collection activities, and must seek the consent of respondents. More specifically, when seeking personal data through surveys, focus group discussions and interviews, the evaluators should disclose the purpose of the evaluation and with whom the data may be shared. They should also provide contact details of the person/entity to refer to for any concerns on the use of personal data. In the course of the evaluation, personal data on beneficiaries should be encrypted and stored to restrict their accessibility. The IET should establish in advance the data retention plan; that is, how long the personal data collected should be available for use by the evaluation.

Baseline analysis. The IET cleans data and conducts baseline analysis. As there are no programme results at this point, baseline analysis will consist of descriptive statistics. The average values of the demographics of the intervention and comparison groups should be compared to ensure that the necessary similarities exist between the two groups, and any statistically significant differences should be noted.

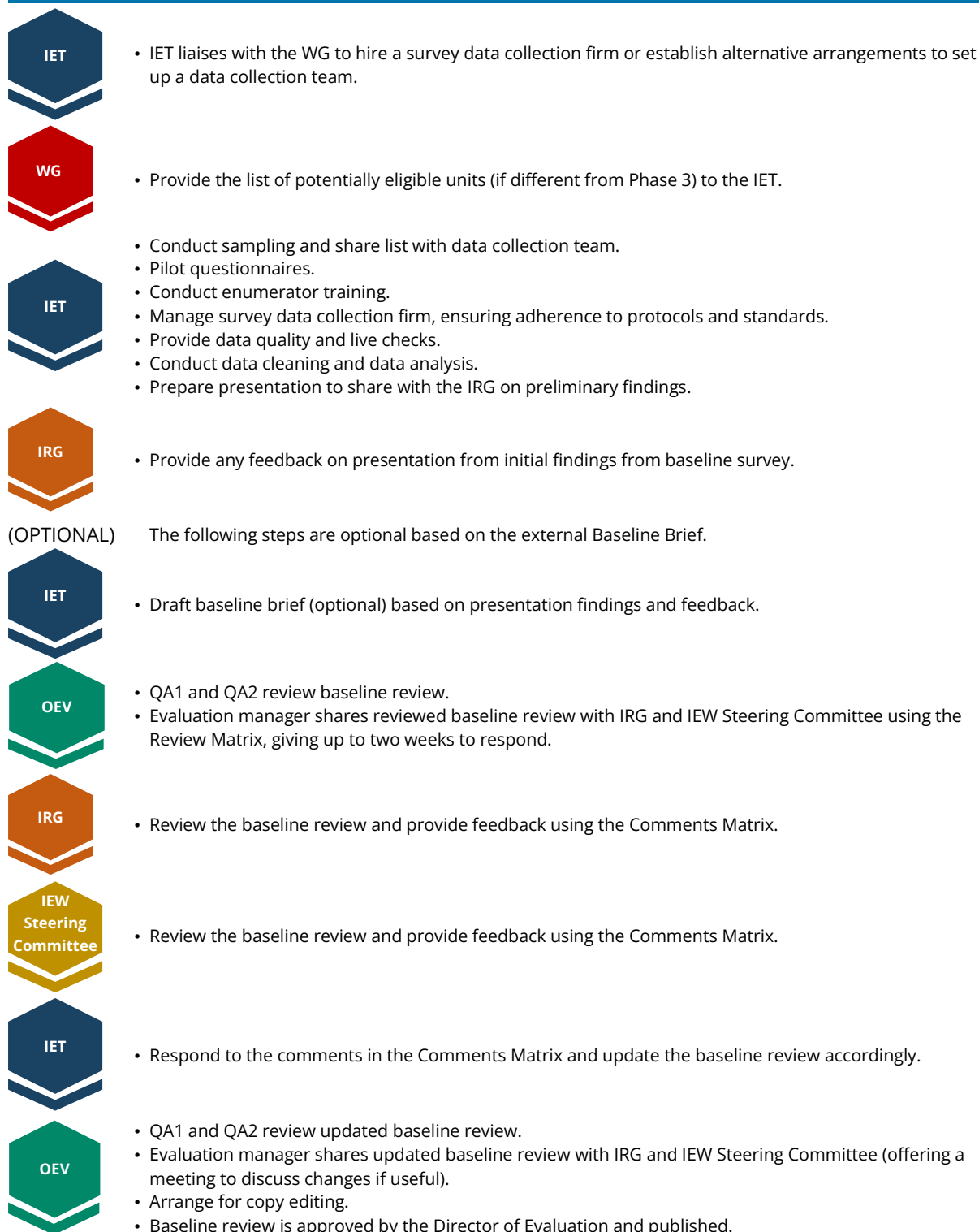
Baseline presentation and discussion of preliminary findings. The IET will present and discuss with the IRG the preliminary findings from the baseline analysis. The content of the presentation and discussion should include, at a minimum:

- background;
- purpose (evaluation objectives and uses);
- overview of evaluation approach and methodology;
- preliminary findings;
- areas for consideration by WFP (but no explicit recommendations);
- any issues that arose during the data collection;
- considerations/concerns over the assumptions in the evaluation design in light of the baseline data; and

- next steps, including feedback loops and opportunities for engagement.

Baseline Brief. In some cases, the baseline analysis can be included in the published Inception Note and annexes. However, depending on the outputs agreed upon at the time of the MoU and Concept Note, the IET can also produce an external baseline analysis brief, which is used for internal and external communications with key stakeholders (e.g. donors).

Figure 9: Process map for the baseline analysis



3.7 BASELINE OUTPUTS

The outputs for the baseline analysis are:

- ✓ baseline survey questionnaires;
- ✓ primary survey data;
- ✓ Stata Do-files (or replicable coding for the analysis);
- ✓ baseline presentation; and
- ✓ baseline brief (optional).

3.8 BASELINE ANALYSIS CONTENT GUIDE AND QUALITY STANDARDS

The purpose of this guidance material is to assist the IET in drafting the externally published Baseline Brief (optional). The Baseline Brief should follow the structure described in the [Baseline Brief template](#). The content should cover the minimum requirements as per the quality checklist.

Data should be presented in a clear and concise manner (in tables, diagrams, etc.) as appropriate for effective communication. They should be systematically analysed and interpreted. Findings should be evidence-based and relevant to the evaluation questions under review. The IET should make a clear distinction between facts borne out by evidence and assumptions or plausible associations they draw from the evidence. Conclusions should follow logically from the analysis of data and findings. The report should be balanced and impartial and use constructive language.

The EM, with the support of the evaluation analyst, carries out first-level quality assurance of the Baseline Brief, using the quality checklist. The EM shares the revised Baseline Brief with the IRG and ERG, if applicable, for comments using the [Comments Matrix](#). The revised Baseline Brief is then reviewed by QA1, QA2 and the Head of the Impact Evaluation Unit (if different from QA2).

3.9 REFERENCE MATERIAL FOR THE BASELINE STUDY

Templates and quality checklists	Technical notes	Other reference material
<ul style="list-style-type: none"> ✓ Comments Matrix 	<ul style="list-style-type: none"> ✓ Technical Note on Principles, Norms and Standards ✓ Technical Note for Planning and Conducting Evaluations During COVID-19 ✓ Technical Note on Stakeholder Analysis ✓ Technical Note on Integrating Gender, Quick Guide and Checklist ✓ Technical Note on Evaluation Matrix 	<ul style="list-style-type: none"> ✓ OEV Communication Protocol ✓ Centralized Evaluation Formatting and Editorial Guidelines ✓ WFP Guide to Personal Data Protection and Privacy

Phase 4. Implementation monitoring and/or high-frequency data collection

4.1 IMPLEMENTATION MONITORING

During the implementation monitoring phase, the country office implements the intervention as agreed in the evaluation design in the selected villages/households/schools. The IET works closely with country office monitoring and evaluation colleagues to ensure that monitoring systems are adequately collecting process and implementation data. The FC supports the country office monitoring and evaluation colleagues and WG by using programme monitoring data to check alignment with agreed targeting and implementation plans.

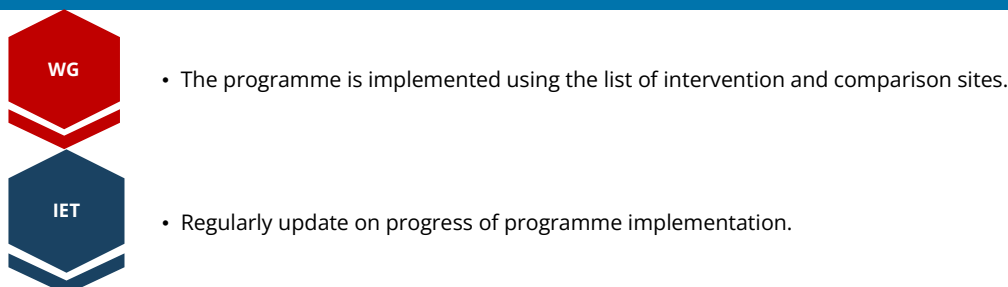
In the case that programme implementation and evaluation design start to diverge (for example, if access is restricted to certain areas, etc.), the IET and WG should escalate to the IRG. Table 6 outlines threats to the evaluation design and risk mitigation strategies.

Table 6: Threats to the evaluation design and risk mitigation strategies

Threats to the evaluation design	Risk mitigation
<p>Loss to follow-up Tracking respondents throughout the evaluation is important because if those surveyed at the baseline cannot be found for the endline survey, it can introduce biases into the analysis and reduce the value of findings.</p>	<p>Gather good contact information during the baseline survey.</p> <p>For evaluations that have a significant length of time between the baseline and endline, such as two years or more, high-frequency data collection when budget allows can reduce loss to follow-up. At a minimum, a short tracking survey can be used to estimate the likely attrition rate and gather additional information.</p>
<p>Attrition refers to the dropout rate of participants or survey respondents. This represents a problem for the evaluation because the dropouts are likely to be systematically different from those who can be found, thus skewing our results.</p> <p>In general, if more than 20% of the respondents leave or do not participate in the survey, it can have a negative impact on data quality. Further, if attrition rates differ by more than ten percentage points across the intervention and comparison groups, it can bias the data in follow-up rounds of data collection.</p> <p>Attrition can occur for any number of reasons, such as loss of interest in the programme, migration or simply unwillingness to participate in the survey.</p>	<p>Motivate participants in intervention and comparison groups to be available for future surveys. This may include providing multiple time slots to the respondent and adjusting according to their convenience, as well as using incentives.</p>
<p>The comparison group is contaminated if it is subject to a similar intervention (external) which affects the outcome of interest or to spill-over effects from the intervention. Spill-over refers to an indirect effect on a subject not directly treated by the experiment which may result in a biased impact estimate.</p>	<p>Spill-over effects cannot be detected accurately in retrospect unless the design considers their existence from the start. Accounting for spillover effects is necessary for correct identification and estimation of direct/intended and indirect/unintended effects. Data should be collected to explain the cause of these effects and who is being affected.</p>

In the case that high-frequency data collection is required, data collection processes will be put in place by the FC, in agreement with the WG, following baseline data collection agreements as outlined in the Concept Note.

Figure 10: Process map for the implementation monitoring phase



4.2 MIDLINE ANALYSIS (OPTIONAL)

While a midline analysis is optional, for longer evaluations (more than four years) it is recommended to include it to assess short-term effects and minimize risks in case of design disruptions. The objective for the midline phase is to produce a Midline Brief through the following steps:

- **Set up data collection team:** The IET will liaise with the WG to hire a data collection firm (likely to be the same as per pilot/baseline) or establish alternative arrangements to set up a data collection team. The IET will then organize enumerator training and set up the data quality system and checks.
- **Midline data collection:** Using the list of sampled villages/households/schools from the randomization phase, the IET will collect midline/high-frequency data.
- **Midline data analysis:** The IET will clean and conduct midline/high-frequency data analysis.
- **Discussion findings:** The IET will present and discuss with the IRG the findings from the midline/high-frequency data analysis.
- **Midline Brief (optional):** The IET can produce a Midline Brief, using the Midline Brief template and the quality assurance process presented below.

4.3 MIDLINE ANALYSIS CONTENT GUIDE AND QUALITY STANDARDS

The purpose of this guidance is to assist the IET in drafting the Midline Brief. The Midline Brief conveys the results of the baseline survey in a way that corresponds to the information needs of the intended users. Evaluation teams have final responsibility for the content of the Midline Brief. The Director of Evaluation has the final authority to approve and publish the report. The evaluation report should follow the [Midline Brief template](#) and should cover the minimum requirements as per the quality checklist.

Data should be presented in a clear and concise manner (in tables, diagrams, etc.) as appropriate for effective communication. They should be systematically analysed and interpreted. Findings should be evidence-based and relevant to the evaluation questions under review. The IET should make a clear distinction between facts borne out by evidence and assumptions or plausible associations they draw from the evidence. Conclusions should follow logically from the analysis of data and findings. The report should be balanced and impartial and use constructive language.

The EM, with the support of the evaluation analyst, carries out first-level quality assurance of the Midline Brief, using the quality checklist. The EM shares the revised Midline Brief with the IRG and ERG, if applicable, for comments using the [Comments Matrix](#). The final Midline Brief is then reviewed by QA2 and the Head of the Impact Evaluation Unit (if different from QA2).

4.4 REFERENCE MATERIALS FOR MIDLINE ANALYSIS

Templates and quality checklists	Technical notes	Other reference material
✓ Comments Matrix	✓ Technical Note on Principles, Norms and Standards ✓ Technical Note on Integrating Gender, Quick Guide and Checklist	✓ OEV Communication Protocol ✓ Centralized Evaluation Formatting and Editorial Guidelines ✓ WFP Guide to Personal Data Protection and Privacy

Phase 5. Endline data collection

The principles of independence, credibility and utility, as well as ethics, have several specific implications at the endline phase, as follows:

Independence: Ensuring independence and impartiality means the non-participation by WFP programme or partner organization staff in the IET's data collection activities with beneficiaries and external stakeholders. The IRG ensures that the IET has full access to available information and data as per the WFP directive on information disclosure. Evaluators have full freedom to conduct their evaluative work without interference or fear for their career. Any challenges are reported to the Director of Evaluation in a timely manner to facilitate resolution.

Credibility: The EM ensures that the evaluation is implemented as per design. Data should be collected in both intervention and comparison groups using the same tools. Importantly, questionnaires should refrain from directly asking about participation in WFP programmes and should be structured to avoid courtesy bias. If challenges arise during the field mission, adjustments are made, ensuring that those do not undermine impartiality.

Utility: The EM organizes a discussion on the preliminary findings at the end of the endline data analysis and the participation of the country office/regional bureau/headquarters, as appropriate.

Ethics: The evaluators behave ethically in all interactions with stakeholders and beneficiaries. The evaluation team establish clear protocols and procedures for dealing with data collection of sensitive and personal data that are in line with the most updated rules and regulations. They ensure informed consent, protecting privacy, confidentiality and anonymity of participants and cultural sensitivity. They respect the autonomy of participants. They ensure the participants' fair recruitment and that evaluation results do not harm them or their communities. They have an obligation to report any noted incidents of fraud, corruption, PSEA or other wrongdoing.¹⁰

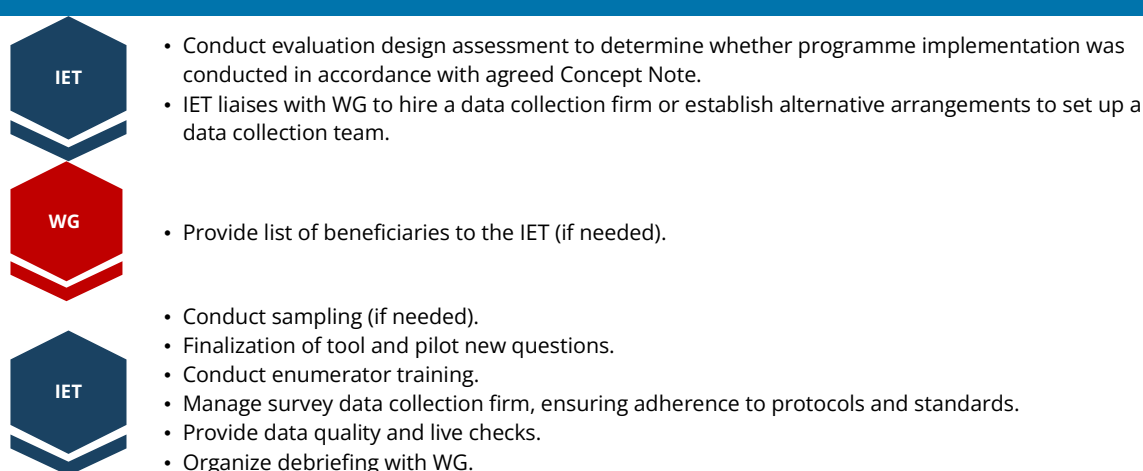
The objective of this phase is to collect endline data through six steps as follows:

- I. **Evaluation design assessment.** Before starting endline data collection, the IET and WG will conduct an evaluation design assessment to confirm that the programme was implemented according to the Concept Note (internal) and that the evaluation design is still relevant and able to produce intended evidence as indicated in the Inception Note (external). At this stage the IET, in coordination with the WG, will determine relevant questions and approaches for the qualitative component.
 - 1 **Finalization of endline tools.** Both qualitative and quantitative data will inform the final report. Depending on the design, the sequencing of qualitative and quantitative data collection can vary, but the different data sources should always complement each other and be presented together in the final evaluation report. New questions will need to be piloted and revised as necessary. If substantial changes are provided to the questionnaire, an amendment to the IRB application will be required.
 - 2 **Set up data collection team.** The IET will liaise with the WG to hire a data collection firm (likely to be the same as per pilot/baseline) or establish alternative arrangements to set up a data collection team. The FC will then organize enumerator training and set up the data quality system and checks. Qualitative data can be collected by members of the IET, depending on the skills of existing members, or by additional qualitative researchers, who can be hired and added to the IET based on needs.
 - 3 **Sampling.** Using the list of all potentially eligible units (e.g. villages/households/schools), the IET will perform sampling on the units to be surveyed (if needed). The sample list will then be shared with the data collection team.

¹⁰ See guidance here: [Where to seek support and report wrongdoing in WFP | WFPgo](#).

- 4 **Endline data collection.** The IET manages endline data collection, ensuring data quality and adherence to protocols and standards.
- 5 **Data protection and confidentiality.** The evaluation team must ensure confidentiality and the protection of data when processing, storing and/or transferring personal data throughout data collection activities, and must seek the consent of respondents. More specifically, when seeking personal data through surveys, focus group discussions and interviews, the evaluators should disclose the purpose of the evaluation and with whom the data may be shared. They should also provide contact details of the person/entity to refer to for any concerns on the use of personal data. In the course of the evaluation, personal data on beneficiaries should be encrypted and stored to restrict their accessibility. The evaluation team should establish in advance the data retention plan; that is, how long the personal data collected should be available for use by the evaluation.

Figure 11: Process map for the data collection phase



5.1 ENDLINE DATA COLLECTION PHASE OUTPUTS

The outputs for the endline data collection phase are:

- ✓ final quantitative survey questionnaires (English and translated);
- ✓ qualitative data collection tools;
- ✓ survey raw data; and
- ✓ qualitative translated transcript.

5.2 ENDLINE DATA COLLECTION PHASE CONTENT GUIDE AND QUALITY STANDARDS

The IET makes a debrief presentation to the WG at the end of the data collection phase to share preliminary findings and conclusions prior to report writing. The debrief presentation is a working document of the evaluation team and will not be commented on or revised. Debriefing with the WG marks the conclusion of the data collection phase. The contents will be guided by the subject and data collected.

5.3 REFERENCE MATERIALS FOR ENDLINE DATA COLLECTION PHASE

Technical notes	Other reference material
<ul style="list-style-type: none"> ✓ Technical Note on Principles, Norms and Standards ✓ Technical Note for Planning and Conducting Evaluations During COVID-19 ✓ Technical Note on Integrating Gender, Quick Guide and Checklist 	<ul style="list-style-type: none"> ✓ OEV Communication Protocol ✓ WFP Guide to Personal Data Protection and Privacy

Phase 6. Analysis and reporting

The principles of independence, impartiality, credibility and utility, as well as ethics, have a number of specific implications at the endline and final report phase:

Independence: The evaluation team must be free from pressure to alter conclusions and recommendations in any way that is not supported by the evaluation's findings.

Impartiality: The evaluation team should analyse data and present findings transparently according to the Concept Note and PaPs. The evaluation team should explicitly reflect where findings are not consistent or deviate from what was indicated in the IR. The report should benefit from inputs from key stakeholders and a rigorous quality assurance process, including external anonymous peer reviewers.

Credibility: This requires that findings are based on PaPs and rigorous evaluation designs. Findings and conclusions must be fair and acknowledge the existence of differing evidence strengths. The evaluation report explains the methodology and any limitations, and presents evidence, findings, conclusions and recommendations in a complete and balanced way.

Utility: A stakeholder workshop should be organized to present initial findings, conclusions and recommendations to the stakeholders.

Ethics: The EM should inform the Head of the Impact Evaluation Unit and Director of Evaluation if there are allegations of wrongdoing and misconduct without breaking confidentiality.

The objective of the report phase is to conduct the analysis, discuss preliminary findings with relevant stakeholders and produce the final evaluation report. The main output of the reporting phase is the final evaluation report.

The reporting phase is undertaken after the data collection phase in order to synthesize, analyse, validate and interpret all data collected. The IET will conduct data analysis according to what is reported in the IR and PaP. Any deviation should be properly discussed and motivated. The EM has primary responsibility for timely delivery of the evaluation report using the following evaluation report template, which has been quality assured internally prior to submission and meets Office of Evaluation quality standards. In addition, there are various levels of quality assurance by the Office of Evaluation as outlined in section 8.3.

Figure 12: Process map for the final report phase



Stakeholder workshop (in-person or virtual): The IET will present and discuss the findings from the final report with the IRG (and ERG where relevant). Prior to the discussion, a draft report should be shared with a request for comments to be submitted (ideally prior to the discussion). The content of the presentation should include, at a minimum:

- background;
- purpose (evaluation objectives and uses);

- overview of evaluation approach and methodology;
- preliminary findings and conclusions; and
- next steps, including feedback loops and opportunities for engagement.

The presentation should be followed by a discussion on the findings and recommendations. Feedback provided by key stakeholders on the preliminary findings should be used to inform the formulation of recommendations for the endline study report.

6.1 FINAL EVALUATION PHASE OUTPUTS

The outputs for the final evaluation report phase are:

- ✓ Stata Do-files (or replicable coding for the analysis);
- ✓ cleaned data (quantitative);
- ✓ coded data (qualitative); and
- ✓ final evaluation report.

6.2 FINAL EVALUATION REPORT PHASE CONTENTS GUIDE AND QUALITY STANDARDS

The purpose of this guidance is to assist the IET in drafting the evaluation report and conveying the results of the evaluation in a way that corresponds to the information needs of the intended users and answers the evaluation questions and related sub-questions. The IET has final responsibility for the content of the evaluation report. The Director of Evaluation has the final authority to approve and publish the report.

Quantitative and qualitative data should be presented in a clear and concise manner (in tables, diagrams, etc.) as appropriate for effective communication. They should be systematically analysed and interpreted. Findings should be evidence-based and relevant to the evaluation questions under review. The evaluators should make a clear distinction between facts borne out by evidence and assumptions or plausible associations they draw from the evidence. Conclusions should follow logically from the analysis of data and findings. The report should be balanced and impartial and should use constructive language. Recommendations should be no more than ten (preferably fewer) and should be relevant, realistic (implementable), prioritized and sequenced.

The evaluation report should specifically consider gender, equity and inclusion issues. This implies ensuring that the analysis, findings, conclusions and considerations for future programmes adequately cover equality and empowerment of women. In addition, a number of ethical considerations should be reviewed at the reporting stage (see Annex A of the *UNEG Ethical Guidelines for Evaluation*).

The evaluation report should follow the Office of Evaluation template and should cover the minimum requirements as per the quality checklist.

In order to minimize formatting issues, the evaluation team should adhere to the template and content guide. The EM is responsible for providing a report and annexes in compliance with the Office of Evaluation’s editorial and formatting guidelines.

The EM carries out first-level quality assurance of the final report, using the quality checklist. The EM submits the evaluation report for second-level quality assurance and clearance. The Head of the Impact Evaluation Unit (if different from QA2) shares with the IRG for comments using the [Review Matrix](#). The updated evaluation report is then reviewed by QA2 and the Head of the Impact Evaluation Unit (if different from QA2) and sent to the Director of Evaluation for approval.

6.3 REFERENCE MATERIALS FOR THE REPORT PHASE

Templates and quality checklists	Technical notes	Other reference material
<ul style="list-style-type: none"> ✓ Template ✓ Example Evaluation report ✓ Comments Matrix template 	<ul style="list-style-type: none"> ✓ Technical Note on Principles, Norms and Standards ✓ Technical Note on Integrating Gender, Quick Guide and Checklist 	<ul style="list-style-type: none"> ✓ OEV Communication Protocol

Phase 7. Dissemination and learning

The principles of independence and utility have a number of specific implications in the dissemination and learning phase as follows:

Independence: All final evaluation reports and post-hoc quality assessment results are published on WFP websites and disseminated through various channels.

Utility: All impact evaluations detail considerations for future programmes. Opportunities for wider organizational and sectoral learning are pursued, including arranging discussions on evaluation results in key workshops and conferences.

Learning and dissemination take place throughout the entire impact evaluation process. At each phase (e.g. planning, preparation, baseline analysis, midline/high-frequency analysis, etc.) the IET should present data, findings or lessons learned to relevant stakeholders in-country and globally through targeted meetings and events. In many countries where WFP operates, access to quality data is limited, and baseline surveys can also be used to inform governments and partners about the situation in communities targeted and any implications for programming.

It is important that evaluation reports are accessible to a wide audience, as foreseen in the WFP Evaluation Policy, to ensure the credibility and utility of evaluations through transparency. Dissemination of the reports and evidence should follow the communication and dissemination plan designed at the inception phase of the evaluation, including a webinar/workshop to present and discuss the findings with the key stakeholders of the evaluation, as needed.

This section provides an overview of the final steps in the evaluation process to ensure that evaluations are accessible to the audience of WFP. Specifically, this phase consists of the following components:

- editing and design of full evaluation report;
- dissemination of evaluation products;
- archiving of data/information related to the evaluation; and
- administrative completion.

7.1 FOLLOW-UP AND DISSEMINATION PHASE PROCESS GUIDE

Detailed guidance and communication options can be found in the communication and knowledge management plan. This section also covers some of the final administrative issues that must be addressed by the EM.

Report editing, formatting and web publishing. The Evaluation Policy specifies that full evaluation reports are public documents available on WFPgo and the WFP.org 'Impact Evaluation' web page.

Evaluation report and dissemination process:

- Send the full evaluation report to the Office of Evaluation Communications and Knowledge Management Unit for editing and final formatting as per corporate/Office of Evaluation standards as soon as the evaluation report has been approved by the Director of Evaluation. Reports will also be translated into the official language of the country engaged in the impact evaluation when needed (to be determined by the Office of Evaluation in discussion with the country office).
- Coordinate with the Office of Evaluation Communications and Knowledge Management Unit for development of dissemination products to accompany the posting of the final full evaluation report on the internet and intranet. Dissemination products might include infographics, videos and podcasts building on the findings in the report.

- Submit the edited and formatted evaluation report and dissemination products for 'final approval' by the Director of Evaluation.
- Alert the post-hoc quality assessment coordinator that the report is ready for post-hoc quality assurance.
- Draft and clear with the Director of Evaluation an introductory paragraph to the evaluation report for the WFP.org 'Independent evaluation' web page. This paragraph should not exceed 600 characters and should include a breakdown and categorization of main findings (which should not exceed seven categories).
- Review the final edited and formatted version of the evaluation report and, when satisfactory, share it with the Office of Evaluation Communications and Knowledge Management Unit. They will review and publish the report and the introduction on WFPgo and the WFP.org 'Independent evaluation' web page and create the required links to topics and countries.

Dissemination of evaluation reports and products. The evaluation report should be disseminated actively with stakeholders throughout the organization. In addition to the report, the IET will prepare blogs, briefs and other relevant communication products to disseminate findings. The IET will consider which evaluations under an IEW will be eligible for submission to an academic journal for publication.

The EM, supported by the research analyst, is responsible for:

- requesting that the Office of Evaluation Communications and Knowledge Management Unit review and publish the evaluation brief on WFPgo and the WFP.org 'Independent evaluation' web page;
- drafting an email to be sent out by the Director of Evaluation to share the final version of the report with the IRG, ERG and other key stakeholders; and
- ensuring implementation of the communication and knowledge management plan developed for the evaluation.

Archiving of closed evaluations. Through the evaluation process, a wide range of formal and informal outputs are created, including documents, data and communications. Such products are an integral part of the evaluation process and should therefore be retained for future reference – for transparency, accountability and internal learning purposes. The Office of Evaluation's Evaluation Management Information System on SharePoint facilitates this.

The EM, with support from the research analyst, is responsible for:

- selecting files for inclusion in the system, which includes all the deliverables in each phase (e.g. questionnaires, Stata Do-files, anonymized data, reports, etc.); and
- delivering a fully archived evaluation, including the reference library, at the end of the evaluation cycle.

The Office of Evaluation's Evaluation Management Information System guidelines give details on the filing/archiving process, file structures, and roles and responsibilities.

Finalization of administrative matters. Within one month of the finalization of the evaluation report, the EM should:

- in cases where individual consultants have been hired, finalize with the Office of Evaluation's business support associate any outstanding payments by reviewing the status of travel expense claims and payments (to consultants as per attendance sheet or firms as per invoices);
- advise the Head of the Impact Evaluation Unit to release uncommitted funds (if any);
- in cases where individual consultants/staff have been hired, finalize the performance assessment requirements in the Performance and Competency Enhancement (PACE) system for each consultant hired directly by the Office of Evaluation;
- complete/update Management Information System requirements; and

- ensure that consultants, evaluation partners and data collection firms follow the protocol outlined in the IRB for the treatment of personal data, by removing all identifiable data from laptops and returning all WFP equipment.

7.2 REFERENCE MATERIALS FOR THE FOLLOW-UP AND DISSEMINATION PHASE

Templates and quality checklists	Other reference material
<ul style="list-style-type: none"> ✓ Template / word guidance for 4-page brief ✓ Example brief 	<ul style="list-style-type: none"> ✓ OEV Communication Protocol

Annex I: Ethics in impact evaluations

This section will present some of the practices designed to ensure ethical impact evaluations. Ethics are a key component in all impact evaluations, studies and research in general.

These ethical considerations and practices are important for a number of reasons. First, to ensure that none of the practices in our work might create any harm or risk to the people we work with. Second, to minimize reputational risks which might affect WFP's reputation if evaluations and studies are not conducted in a way that is ethical. Finally, to ensure greater quality and credibility: respondents who feel treated well and with respect are less likely to respond in ways that would compromise the validity of the information collected. Cooperating partners are also more likely to cooperate and not withhold important information.

All impact evaluations directly involving human subjects or the use of **personally identifiable information** must identify and secure ethics approval from an **IRB**. IRB approval typically takes place after survey instruments are finalized, but before any intervention is implemented or data collected. Approval is also required before any changes to study design. Finally, approval from national institutions is also important, as it will ensure that a study is compliant with the laws and social norms of the country where the data originate.

In addition to obtaining IRB approval for any impact evaluation, IET members who handle personally identifiable information must have up-to-date Human Subjects Research Certification, which ensures that they are aware and knowledgeable of practices to protect evaluation participants.

The IRB will review and monitor the evaluation design, questionnaires and protocols to ensure that they protect the rights and welfare of human subject research participants. While recognizing the existence of comparable regulations, the IRB typically follows three ethical principles reported in the US [Belmont Report](#), which regulates the Federal (US) Policy for the Protection of Human Subjects (the "Common Rule"), outlines criteria for IRB review and establishes minimum ethical principles. These are the principles of **Respect for Persons, Beneficence** and **Justice**.

RESPECT FOR PERSONS

The **Respect for Persons** principle states that individuals should be treated as autonomous agents. This means recognizing individuals' rights and abilities to decide for themselves what to do, what to participate in, and what to allow their personal information to be used for.

In practice, this principle requires:

- 1 seeking individuals' **informed consent**;
- 2 establishing a form of **compensation**; and
- 3 additional protections for individuals who are considered vulnerable.

Seeking **informed consent** means ensuring that consent (or assent in the case of minors) is truly clear and easy to understand, allowing individuals to independently decide to participate in a study. This is a pre-condition/requirement before conducting any survey. At the very minimum, all surveys will start with an enumerator reading a form to ensure that respondents are informed about the implications of joining a study. It is good practice to also leave the respondent with a written form, with the information of who to contact if they withdraw their consent during a later stage. In many cases, communication with local authorities and leaders is also required to inform participants.

The consent form needs to include:

- the purpose of the evaluation, using a language that can easily be understood, and why the respondent was selected;
- what their participation entails (e.g. duration of the interview);
- confirmation that participation is voluntary and confirmation that consent can be revoked at any time;
- any foreseeable benefits or risks and any form of compensation for participating;

- a description of how identity will be protected and what information will be disclosed and to whom; and
- details of a local contact person for questions/concerns/withdrawal.

Compensation should offset the time and inconvenience of participation. It is meant as a benefit of research and is **in addition to** reimbursement of direct expenses. It could possibly be in the form of in-kind gifts. It should not be too low or high as to potentially undermine participants' abilities to make a rational decision about participating in the study.

Finally, the Respect for Persons principle also recognizes that some people – such as minors, prisoners and those who are otherwise vulnerable – may have diminished autonomy and that providing information may not be enough to allow them to make an informed decision in their best interest about their participation in research. In these cases, **additional protections are required**.

BENEFICENCE

The **Beneficence** principle includes two main concepts:

- 1 the “do not harm” principle; and
- 2 the “maximize benefits while minimizing risk” principle.

The **“do not harm”** principle requires researchers to evaluate for each study if the future benefits from the evaluation/research justify the risk to the subjects. The “do not harm” principle has two primary implications.

First, do not deny participants an intervention, treatment or services they would otherwise be entitled to receive. In practice, in the context of an impact evaluation, this principle translates as there being opportunities for randomization when:

- there is limited implementation capacity and a programme must be implemented in steps;
- budget constraints prevent full coverage of all those who could be eligible; and
- you are piloting a new way of implementation and there is no evidence on what works best.

Second, do not administer an intervention or treatment or conduct data collection that is known or highly likely to be harmful. For example, if there is evidence that an intervention is posing harm to a particular population, the experiment and intervention need to be interrupted immediately.

The second principle under Beneficence is **“maximize benefits while minimizing risks”**. This principle requires that potential benefits are evaluated in the context of the credibility of the evaluation results. It means, for example, that a poorly designed study where the results are not credible has no benefit and would therefore fail a risk-to-benefit ratio assessment because it would expose individuals to the burdens of participation without any knowledge gains.

Some practices to maximize benefits during impact evaluations include widely disseminating evaluation findings, ensuring findings are used to improve programmes and policies, sharing findings and information with local communities, and working with key stakeholders to ensure action is taken as a result of the findings.

Aside from any risks from the intervention in a randomized study, risks in evaluations/research also arise from the data collection itself – for example, risks in exposure of sensitive or private data or in how data are collected. For this reason, it is essential that protocols are designed to minimize potential harms. Some examples of protocols and practices include:

- protocols and procedures for the data life cycle (including collection, transfer, process, publication and destruction);
- interview protocols, with clear instructions on where and how to conduct interviews (ideally in private whenever possible and contextually appropriate);
- construction of questions which are sensitive;

- developing a protocol for identifying when people are overly distressed, and reporting and escalation plans for when this occurs; and
- how to train staff on reacting appropriately to respondent answers, especially if the topics are potentially sensitive or emotional.

Pilot interviews with field staff and in focus groups with out-of-sample community members to make sure questions are appropriate to ask and worded appropriately. Consider the cognitive burden of questions and the time required for the study. Factor in how the location of the study affects the burden to participants. Finally, develop communication protocols for cooperating partners to ensure consistent and adequate messages with communities involved in the study and to set the right level of expectations.

JUSTICE

Finally, the **Justice** principle requires that the benefits of research do not go only to one group and the burdens to another. For each intervention arm of the impact evaluation, there must be a genuine lack of evidence about how one kind of intervention compares (e.g. is more or less effective) with the other interventions arms, as well as the best viable alternatives, **Box 2: Principles of UNEG's Ethical Guidelines for Evaluation**

UNEG developed its *Ethical Guidelines for Evaluation* to ensure that an ethical lens informs day-to-day evaluation practices. While there is no overall agreed definition of "ethics" for the UN system, the guidelines define ethics as "the right or agreed principles and values that govern the behaviour of an individual within the specific, culturally defined context within which an evaluation is commissioned or undertaken". Four ethical principles are identified:

Integrity is the active adherence to moral values and professional standards, which are essential for responsible evaluation practice. Integrity in evaluation requires: (1) honesty and truthfulness in communication and actions; (2) professionalism based on competence, commitment, ongoing reflective practice and credible and trustworthy behaviour; and (3) independence, impartiality and incorruptibility.

Accountability is the obligation to be answerable for all decisions made and actions taken; to be responsible for honouring commitments, without qualification or exception; and to report potential or actual harms observed through the appropriate channels. Accountability in evaluation requires: (1) transparency regarding the evaluation purpose and actions taken; (2) responsiveness as questions or events arise, adapting intentions and plans as required (for example, where corruption, fraud, sexual exploitation or abuse or other misconduct or waste of resources is identified, it must be referred to appropriate channels); (3) taking responsibility for meeting the evaluation purpose and for actions taken, for exercising due care and for ensuring redress and recognition as needed; and (4) justifying and fairly and accurately reporting to stakeholders (including affected people) decisions, actions and intentions.

RESPECT involves engaging with all stakeholders of an evaluation in a way that honours their dignity, well-being and personal agency while being responsive to their sex, gender, race, language, country of origin, LGBTQ status, age, background, religion, ethnicity and ability, and to cultural, economic and physical environments. Respect in evaluation requires: (1) access to the evaluation process and products by all relevant stakeholders; (2) meaningful engagement and fair treatment of all relevant stakeholders in the evaluation process from design to dissemination, so they can actively inform the evaluation approach and products rather than being solely a subject of data collection; and (3) fair representation of different voices and perspectives in evaluation products.

BENEFICENCE means striving to do good for people and the planet while minimizing harms arising from evaluation as an intervention. Beneficence in evaluation requires: (1) explicit and ongoing consideration of risks and benefits from evaluation processes, products and longer-term consequences; (2) maximizing benefits at systemic (including environmental), organizational and programmatic levels; (3) doing no harm and not proceeding with an evaluation when harms cannot be mitigated; and (4) ensuring evaluation makes an overall positive contribution to human and natural systems and to the mission of the United Nations.

Annex II: Roles and responsibilities matrix

Steps and responsibilities	Internal Reference Group (IRG)/ Working Group (WG)**					Impact Evaluation Team (IET)					External Reference Group (ERG) (optional)			Window-level governance groups		OEV
	CD/D/CD	Head of Programme	Implementers (country office programme team/CPs)**	Country office M&E (RAM/VAM)**	Regional evaluation officer	Head of Impact Evaluation	Evaluation manager (typically WC)	Impact evaluation analyst	Field coordinator (FC)	External evaluation partners (typically includes technical lead)	External stakeholders (government, other UN agencies, other donors)	Headquarters/regional technical units	Donor	IEW Steering Committee (WSC)	Window Reference Group	Director of Evaluation
Phase 1: Planning																
Initiate request to OEV for impact evaluation	A	L	P	P	S											
Conduct impact evaluation feasibility assessment		P	P	P	P	I	L	S								
Identify evaluation partners/evaluation capacity		I	I	I	I	I/S	L	S		P						
Formal inclusion into window	I	I	I	I	I	S	L	S		P			S			A
Phase 2: Preparation																
Set up IET							L	S		P						
Establish WG, IRG and ERG (optional)	P	P	P	P	P	P	L	S		S	P	P	P			
Hire FC				L/S			S			S/L						
Impact evaluation design workshop		P	P	P	P		L	P	P	P						
Prepare a proposed Concept Note (internal), including evaluation design, data collection tools and Risk Matrix.							P	S		L						
Obtain ethical clearance from IRB							P	S		L						
Register PaP							P	S		L						

Sign MoU with attached Concept Note (internal)	A	S						L			P							A
Provide a list of potentially eligible intervention units (e.g. villages/households/schools)				L				S	S		S							
Randomly select intervention and comparison units (e.g. villages/households/schools) to be involved in the intervention								S	S		L							
Check the list of intervention and comparison units is operationally feasible				L				S	S		S							
Phase 3: Inception																		
Draft Inception Note (external)								P	S		L							
Quality assurance process		P	P	P	P			L	S		P							
Publish Inception Note								L	P		P							A
Baseline analysis (optional)																		
Hire survey data collection firm/set up data collection team				S							L							
Provide list of potentially eligible units				L							P							
Conduct sampling								P	S		L							
Develop and pilot questionnaires								P	S		L							
Train enumerators, supervise and manage data collection, conduct data quality assurance										L	P							
Data cleaning and analysis											L							
Draft Baseline Brief (optional)								P	S		L							
Quality assurance process for Baseline Brief (optional)		P	P	P	P			L	S		P							
Publish Baseline Brief (optional)								L	S		P							A
Phase 4: Implementation and monitoring																		
Implementation of intervention/programme as outlined in the impact evaluation design			L	I				I		I	I							
Regularly update IET and WG			P	P				P		L	P							
Programme implementation monitoring data			P	L				I		P	I							
Phase 5: Endline data collection																		
Hire survey data collection firm/set up data collection team				S							L							
Provide list of potentially eligible units				L							P							
Conduct sampling								P	S		L							
Develop and pilot questionnaires								P	S		L							
Train enumerators, supervise and manage data collection, conduct data quality assurance										L	P							

Phase 6: Analysis and reporting																
Data cleaning and analysis										L						
Stakeholder workshop		P	P	P	P		L	P	P	P						
Draft final report							P	S		L						
Quality assurance process final report		P	P	P	P		L	S		P						
Publish final report							L	S		P						A
Phase 7: Dissemination and learning																
Develop communication products (Blogs, Opt-out)							L/P	L/P		L/P						
Outreach events related to impact evaluation products		P	P	P	P		L	P	P	P	P	P	P		I	

Note: A - approve; L - lead; P - participate actively; S - support; I - be informed

Acronyms

CD	country director
CO	country office
DCD	deputy country director
DOE	Director of Evaluation
EM	evaluation manager
ERG	External Reference Group
HQ	headquarters
IET	Impact Evaluation Team
IEQAS	Impact Evaluation Quality Assurance System
IRB	Institutional Review Board
IRG	Internal Reference Group
M&E	monitoring and evaluation
MIS	Management Information System
OEV	Office of Evaluation
PaP	pre-analysis plan
PSEA	Protection from Sexual Exploitation and Abuse
RAM	research, assessment and monitoring
RB	regional bureau
SDG	Sustainable Development Goal
TL	technical lead
TOR	terms of reference
UNEG	United Nations Evaluation Group
VAM	vulnerability analysis and mapping
WSC	Window Steering Committee
WG	working group

Office of Evaluation

World Food Programme

Via Cesare Giulio Viola 68/70,
00148 Rome, Italy - T +39 06 65131

wfp.org/independent-evaluation