

UNEG
AGM 2025



UNEG
United Nations Evaluation Group

UNEG Evaluation Synthesis Interactive Guidance

UNEG AGM 2025: Background document

This document has been prepared by the Evaluation Synthesis Working Group and is being circulated at the AGM 2025 for information.

The WG will finalise the guidance document in Q1/ Q2 of 2025, including editing before it is circulated to UNEG Heads electronically for approval as a UNEG guidance document.

Contents

1. Introduction	1
2. What is evaluation synthesis?	2
2.1 Definition	2
2.2 Purpose and types of evaluation synthesis	4
2.3 The value of evaluation synthesis	5
2.4 When to conduct an evaluation synthesis	6
3. Decision roadmap	7
4. Governance of a synthesis exercise	8
4.1. Management Group	8
4.2. Advisory group	8
4.3 What are the key considerations for engaging advisory groups?	9
4.4 Can the composition of advisory groups vary for different synthesis types?	10
5. Developing a Terms of Reference for commissioning a synthesis	12
5.1. What is the purpose of a Terms of Reference?	12
5.2. What should a Terms of Reference include?	12
6. Defining evaluation synthesis questions	16
6.1. How can synthesis questions be developed?	16
6.2. What kinds of questions can evaluation synthesis answer?	17
7. Mapping or scoping the evaluative evidence base	21
7.1. What is the purpose of mapping or scoping the evaluative evidence base?	21
7.2. When in the synthesis process can a scoping occur?	21
7.3. How to undertake a mapping or scoping	22
8. Protocol/inception reports	24
8.1 What is the purpose of an inception report or protocol?	24
8.2 Should inception reports or protocols be published?	24
8.3 What should be included in an inception report or protocol?	24
9. Eligibility criteria	26
9.1 What is the purpose of eligibility criteria?	26
9.2 What frameworks are available for developing eligibility criteria?	26
9.3 Examples of eligibility criteria	28
10. Systematic searching	34

10.1	What is the purpose of systematic searching?	34
10.1.	How to develop a search strategy?.....	34
10.2.	What tools can be used to manage references?	36
11.	Screening	37
11.1.	What is the purpose of screening	37
11.2.	How to conduct a screening	37
11.3.	Tools available for screening.....	38
12.	Purposive sampling for qualitative evaluation synthesis	39
12.1.	What is the purpose of purposive sampling in a synthesis?.....	39
12.2.	When does purposive sampling occur in the synthesis process?.....	39
12.3.	What are the key considerations?	40
12.4.	What are the different approaches to purposive sampling?.....	40
13.	Quality appraisal	43
13.1.	What is the purpose of quality appraisal in a synthesis?.....	43
13.2.	How is the quality of evaluations and research studies appraised?.....	43
13.3.	When does quality appraisal occur in the synthesis process?.....	43
13.4.	What are the key considerations?	43
13.5.	Quality appraisal for single agency syntheses	44
13.6.	Quality appraisal for multi-agency syntheses.....	44
13.7.	Quality appraisal in syntheses of UN evaluations and other types of evidence.....	45
14.	Coding and extracting data	48
14.1.	What is data extraction?	48
14.2.	Qualitative data extraction	48
14.3.	Quantitative data extraction.....	48
14.4.	Data extraction forms	49
15.	Analyzing and synthesizing	51
14.1	What is the purpose of analyzing and synthesizing?	51
14.2	Structuring the analysis.....	51
14.3	Overview of synthesis methods.....	51
14.4	Synthesis methods to answer ‘what works’ type questions.....	52
14.5	Qualitative synthesis methods to answer ‘how and why’ type questions	54
14.6	Mixed method and realist synthesis to explore ‘what works for whom and in what circumstances’	55

16. Reporting the results of a synthesis	59
16.1. What is the purpose of reporting in a synthesis?	59
16.2. What are the key considerations for the reporting phase?.....	59
16.3. How could a synthesis report be structured?	60
Annex 1 List of UN Agency evaluation databases	64
Annex 2 List of databases for external research	66

1. Introduction

Members of the UN Evaluation Group have been synthesizing evaluative evidence in various ways for many years. However, pressure on evaluation offices to offer comprehensive and integrated evaluative evidence, understand and demonstrate what UN agencies are achieving, and to assess trends across evaluations to further the SDGs has increased the attention on the value of evaluation synthesis.

The UNEG Synthesis Working Group was created in 2022 as evaluation synthesis was identified as an important area of practice that lacked sufficient coordination and guidance. The group mapped the various synthesis exercises being carried out by individual agencies, as well as inter-agency evaluation synthesis work. While there are common approaches and practices used across agencies, there is also considerable variety in purpose, terminology and process steps in evaluation syntheses across UNEG member organizations.

Inter-agency evaluations focusing on SDG targets, as well as the Global SDG Synthesis Coalition initiative launched in 2022, have broadened the synthesis approaches taken to include evidence from impact evaluations, research studies and other assessments (e.g. Voluntary National Reviews). These examples have expanded the complexity of work being done by UNEG members and have brought evaluation synthesis closer to evidence synthesis (i.e. research) practices in many ways.

This guidance is meant primarily to support the synthesis work of UN Evaluation Group members, which is commonly focused on synthesizing evaluations that have used mixed methods and are increasingly conducting impact evaluation. The use of qualitative, quantitative, mixed methods, and experimental designs enhances the need for syntheses to bring together evidence across these evaluation disciplines.

A **common definition** for evaluation synthesis for the UN system is offered and **minimum standards** for high quality, credible evaluation syntheses – be it agency-specific and focused only on internal evaluations, inter-agency syntheses or syntheses that include qualitative and/or quantitative evaluation evidence alongside evidence from other sources – are highlighted. The guidance works to differentiate process steps and considerations across a range of synthesis approaches.

The document is intended to be a living document and will be adapted to an online format to be interactive and regularly updated. In its current form, it is necessarily linear; the early presentation of a decision-roadmap is meant to help the reader advance to sections that are specific to the type of synthesis that they will carry out rather than having to go through the document in its entirety.

2. What is evaluation synthesis?

2.1 Definition

An evaluation synthesis can be described as the **systematic collation and analysis of existing quality-assessed evaluation evidence** (e.g. findings, conclusions, lessons, recommendations) **by type of evaluation, theme, SDG, or type of intervention to contribute evidence and develop new findings to inform strategic, policy and programme decision-making at different levels.**

The defining characteristics of an evaluation synthesis are:

- **The units of analysis are evaluations.** These are the core activities of UN evaluation offices' work, and they provide systematic, independent and impartial assessments of initiatives, projects, programmes, strategies, policies, topics, themes, sectors, operational areas or institutional performance.¹ In some syntheses, other types of evidence published by the UN and external to the UN may be included.
- **They follow a systematic and transparent approach** to identifying, collating and appraising the quality of individual evaluations, and then synthesizing findings and lessons from bodies of evaluative evidence.

Table 1 below presents the various definitions of synthesis-related terminology used in the UN system. This expands on a mapping exercise of all synthesis-related exercises undertaken by UNEG members, carried out by the UNEG Synthesis Working Group in 2022. This revealed inconsistencies in terminology with agencies using different terms for similar exercises and/or terminology for different exercises. To address this, the Working Group agreed to propose standard definitions for synthesis-related terms within UNEG's work – specifically, evaluation synthesis as it differs from evidence synthesis, evidence gap maps, meta-evaluation, summaries of evaluative evidence, systematic reviews, and rapid evidence assessments.

Table 1: Definitions of synthesis-related terminology used by UN agencies

Exercise	Definition	Indicative timeframe	Minimum requirements	Purpose	
				Accountability	Learning
Evidence synthesis	Evidence synthesis can be described as the process of bringing together information from a range of sources to inform debates and decisions on specific issues. ² This can be useful to identify gaps in knowledge, establish an evidence base for best-practice guidance, or help inform policymakers and practitioners.	9-12 months	A systematic, explicit, transparent and reproducible approach to collating, quality appraising, analyzing and	√	√

¹ For UNEG's full definition of an evaluation, see: United Nations Evaluation Group (2016). Norms and Standards for Evaluation. New York: UNEG. <https://www.uneval.org/document/detail/1914>

² Royal Society. (2018) Evidence Synthesis for Policy: A statement of principles. London: Royal Society/Academy of Medical Sciences. <https://www.royalsociety.org/-/media/policy/projects/evidence-synthesis/evidence-synthesis-statement-principles.pdf>

			synthesizing evidence.		
Evaluation synthesis	Evaluation synthesis is a specific type of evidence synthesis. It entails the systematic collation and analysis of existing evaluation evidence (e.g. findings, conclusions, lessons, recommendations) by type of evaluation, theme, SDG, or type of intervention to develop new findings to inform policy and programme design.	6-9 months	A systematic, explicit, transparent and reproducible approach to collating, quality appraising, analyzing and synthesizing evaluative evidence.	✓	✓
Evidence gap maps	A visual and interactive map providing an overview of the available evidence on a particular topic, theme or domain. Evidence gap maps do not answer specific synthesis questions but do aim to visualize the distribution of evidence, showing where there are gaps or clusters of evidence across an intervention (rows) and outcome (columns) framework.	6-8 months.	A transparent and reproducible approach to collating, quality appraising, coding and presenting evidence in an easy-to-digest framework.		✓
Meta-evaluation	The evaluation of evaluations, a process that assesses the quality, utility, and relevance of evaluation methods, findings, and reports.	9-12 months	an instrument used to aggregate findings from a series of evaluations.	✓	✓
Rapid evidence assessment	Rapid assessments try to establish knowledge around a specific issue. They are similar to a systematic review but quicker due to: narrower scope, less comprehensive searches, methodological short-cuts (e.g., inclusion of specific research designs such as meta analyses or controlled studies).	3-8 months	Similar to a systematic review, but there is some flexibility in the scope and comprehensiveness of coverage.		✓
Summary of evidence	A document that presents, concisely and clearly, an overview of evidence from a defined set of evaluations on a particular topic. ³	4 months	Consolidation of evidence on a specific theme		✓
Systematic review	A comprehensive and transparent approach to finding, collating, critically appraising and synthesizing the results of all research relevant to a research question to provide new knowledge. ⁴	12-15 months	A systematic, explicit, transparent and reproducible approach to collating, quality appraising, analyzing and		✓

³ WFP. Draft Summary of Evaluation Evidence – Guidance; November, 2023.

⁴ According to Campbell Collaboration a systematic review is an academic research paper, also called a report, that uses a method called 'evidence synthesis' to look for answers to a pre-defined question. The purpose of a systematic review is to sum up the best available research on that specific question. This is done by synthesizing the results of several studies. A systematic review uses transparent procedures to find, evaluate and synthesize the results of relevant research. Procedures are explicitly defined in advance, to ensure that the exercise is transparent and can be replicated. This practice is also designed to minimize bias. Studies included in a review are screened for quality, so that the findings of a large number of studies can be combined. Peer review is a key part of the process; qualified independent researchers review the author's methods and results.

			synthesizing evidence. Systematic reviews aim to be comprehensive in their coverage of evidence on a specific topic, theme or research question.		
--	--	--	--	--	--

2.2 Purpose and types of evaluation synthesis

2.2.1 Purpose

Evaluation syntheses in the UN system can have two main different foundational purposes:

Accountability: Evaluation syntheses that are included in forward-looking evaluation plans submitted by Evaluation Offices to their governing bodies have accountability requirements. These synthesis reports are demand-driven and presented to the governing bodies for their consideration and often include recommendations and management responses. Syntheses also can be used to support accountability to affected populations.

Learning: Evaluation syntheses that are learning-focused are sometimes *ad hoc* and demand-driven rather than resulting from a formal commitment made to a governing body. Learning from synthesis can be used to improve programming, inform strategic decisions and support performance management, but also provide insights to guide future direction.

2.2.2 Types

In addition, there are three types of evaluation synthesis in the UN system:

- **Single agency syntheses:** The synthesis of evaluations from a single UN agency. These are managed or undertaken by a single agency to synthesize their own evaluations. They may synthesize evidence on a series of programmes, strategies, policies, sectors or operational areas. They may also look to synthesize performance and achievements at an organizational level.
- **Multi-agency syntheses:** The synthesis of evaluations from multiple UN agencies. These are managed or undertaken by a group of agencies working together. They may synthesize evidence on thematically or operationally specific programmes, strategies, policies, sectors or operational areas. An important objective of many recent multi-agency syntheses is to synthesize evidence across the UN systems on progress towards specific Sustainable Development Goals (SDGs).⁵
- **Syntheses of UN evaluations and other types of evidence:** These could be single or multi-agency syntheses that triangulate UN evaluation evidence and other sources of evidence published by UN agencies and externally. These other sources of evidence could include impact evaluations, systematic reviews, or other types of research relevant to the synthesis objectives.

⁵ Examples include multi-agency synthesis on SDGs 4.5, 5 and 6 and the wider work of the SDG Coalition.

2.3 The value of evaluation synthesis

Evaluation synthesis is increasingly recognized within the UN system as a valuable tool for evidence-informed decision-making. This recognition is based on several key arguments:

2.3.1 Leveraging the UN's vast evaluative evidence base

As of December 2024, the UNEG database contains over 25,000 records. However, the actual number of evaluations published by individual UN agencies far exceeds this figure. This immense repository of evaluative evidence offers a powerful resource for decision-makers within and beyond the UN system. Synthesizing this evidence enables a more comprehensive and aggregated understanding of problems, findings, and lessons learned from the implementation of projects and programmes. It also highlights critical factors affecting implementation, fostering more effective decision-making.

2.3.2 Using multiple sources of evidence to inform decision making

A fundamental principle of evidence synthesis is drawing from multiple sources of evidence. Basing decisions on a single evaluation or selectively using findings from several evaluations increases the risk of biased or incomplete conclusions. A single evaluation may only offer part of the story, while cherry-picking findings without a systematic approach can skew insights. Evaluation synthesis systematically consolidates evidence from multiple evaluations, increasing the scale, representativeness, and generalizability of findings. This approach enhances the trustworthiness of conclusions while considering contextual factors and multiple perspectives.⁶⁷

2.3.3 Mapping and enhancing accessibility to the evidence base

Evaluation synthesis involves mapping the evaluative evidence base for a specific question, topic, or sector and collating this evidence in one place. This improves accessibility by consolidating evidence from various repositories and locations. It also supports the identification of trends over time, systemic challenges, chronic performance issues, and aggregated evidence against thematic priorities – informing strategic decision-making. Additionally, synthesis helps users design new evaluations by identifying what has already been evaluated, the insights produced, and gaps in evidence. This process informs the thematic focus and lines of inquiry for future evaluations, ensuring a targeted and efficient approach.

2.3.4 Uncovering patterns to inform future decision making

By analyzing evidence over an extended period, evaluation synthesis allows for the identification of recurrent issues related to organizational culture and performance. This insight enables decision-makers to address persistent challenges strategically and allocate attention to areas requiring higher priority.

2.3.5 Promoting cost-efficiency and resource optimization

Synthesizing evidence offers a cost-effective alternative to conducting multiple standalone evaluations. By consolidating a large body of evaluative evidence through a structured and systematic approach, synthesis reduces redundancy and avoids duplicating efforts in areas where sufficient evidence already exists. Similarly, it enables better resource allocation by focusing evaluations on areas with evidence gaps, ensuring efficient use of time and funding.

⁶ Bakrania, S. (2020). *Methodological Briefs on Evidence Synthesis: Brief 1 – Overview*. Florence: UNICEF Innocenti. <https://www.unicef-irc.org/publications/pdf/IRB%202020-01.pdf>

⁷ Oliver Sandy, et al. (2018), 'Approaches to Evidence Synthesis in International Development: A research agenda', *Journal of Development Effectiveness*, 10, 3, 2018, pp. 305–326. <https://doi.org/10.1080/19439342.2018.1478875>

2.4 When to conduct an evaluation synthesis

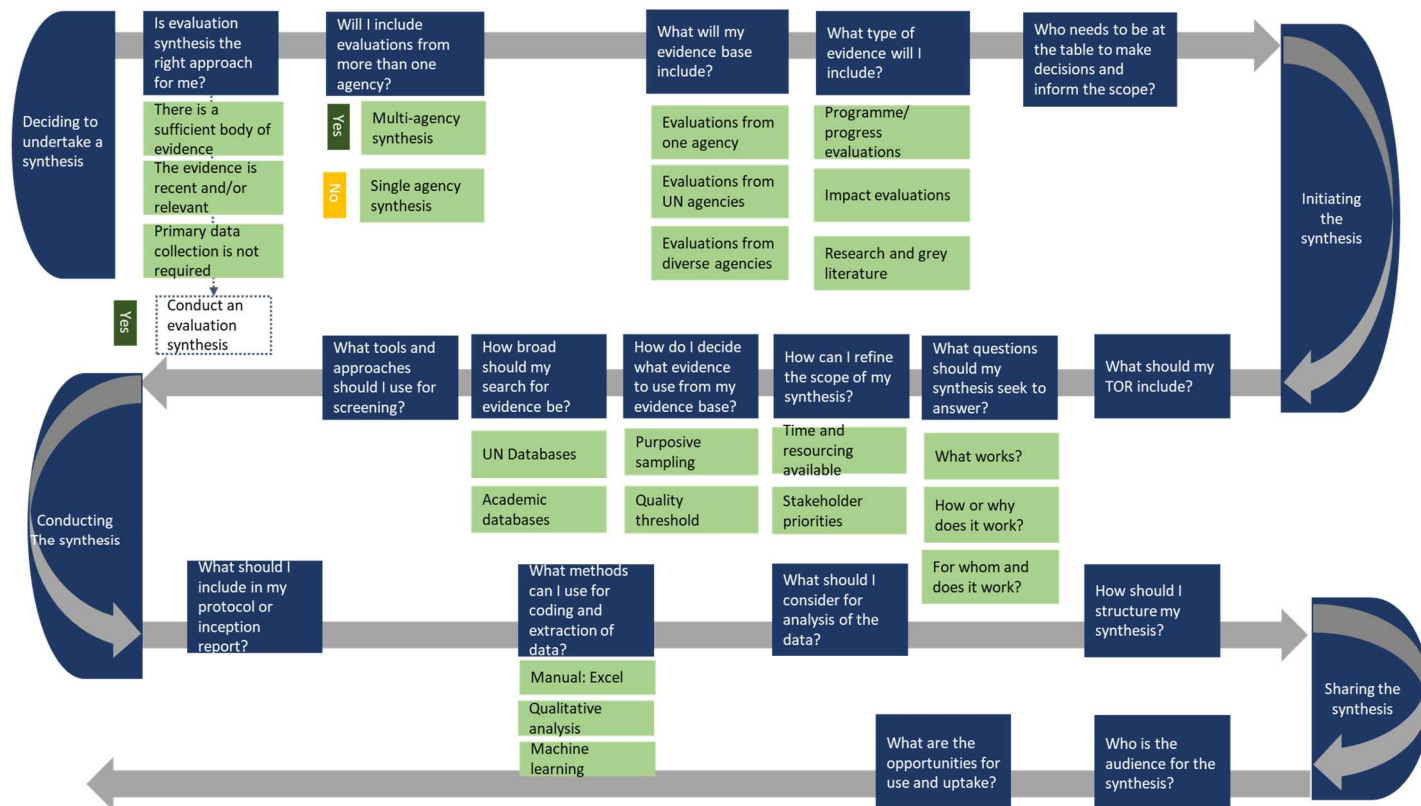
The decision to conduct an evaluation synthesis, rather than an evaluation, is chiefly based upon the availability, recency, relevance, and sufficiency of existing evidence to generate learning and accountability. The considerations for deciding whether to undertake a synthesis are:

- Availability of existing and known evidence: A sufficient body of evaluations have covered a given topic in sufficient depth to be able bring together evidence to identify trends, patterns, or additional findings.
- Recency of the timeframe of the evidence: The evidence available is current and offers meaningful insight for the current context.
- Relevance of the evidence to the users/topic that you are considering: The evidence covers the themes or geography that are of interest to potential users.
- Sufficiency of secondary evidence to answer the proposed research questions: The proposed questions or lines of enquiry can be answered through secondary methods and primary data collection is not needed.

In addition, it is recommendable to carry out evaluation synthesis at regular intervals once evidence becomes available to ensure the evidence synthesis remains up-to-date and continue to provide valuable insights.

3. Decision roadmap

Undertaking a synthesis entails a number of design, governance, and methodological considerations. These decisions are driven by the topic of the synthesis, the evidence available, the questions driving the synthesis, and stakeholders' priorities and interests. This roadmap outlines the key steps and considerations for effectively designing and conducting a synthesis to support decision-making at each stage of the synthesis process.



4. Governance of a synthesis exercise

4.1. Management Group

4.1.1 Purpose

Each synthesis requires a governance structure to provide leadership and coordination of the overall synthesis process. Depending on the size and complexity of the synthesis exercise, there may be one synthesis manager appointed by a UN Evaluation function or a management group comprised of representatives of different agencies. For example, an inter-agency synthesis, such as an SDG synthesis, will be managed collectively by representatives from participating agencies.

Management group agencies typically make a commitment to the synthesis process through financial or human resources and participation in regular management group meetings.

4.1.2 Role

The individual manager or management group is meant to guide and coordinate the process, fulfilling several key roles (these roles can vary):

Planning and design

- Determining synthesis objectives, thematic scope and guiding questions through stakeholder consultation
- Develop synthesis and consultant TORs
- Forming an advisory group
- Convening and participating in advisory group meetings

Conduct

- Quality assuring the overall synthesis process and ensuring that minimum requirements are met
- Providing feedback on key deliverables produced by the synthesis team/consultants (e.g. inception report, interim findings, draft report, evidence gap maps, briefs and presentations)

Communication and update

- Support the communication and dissemination of the synthesis within their agencies and in other relevant forums.

4.2. Advisory group

4.2.1 Purpose

Advisory groups (alternatively known as reference groups or steering groups) are a well-established stakeholder engagement and peer review mechanism used for evaluations and can similarly be utilized for evaluation synthesis.⁸ Inclusive and diverse advisory groups can engage in different stages of the synthesis process to ensure ownership, relevance, credibility, use and uptake of the synthesis.

⁸ See Standard 4.6 of the UNEG Norms and Standards. United Nations Evaluation Group (UNEG) (2016). *UNEG Norms and Standards*. <https://www.unevaluation.org/document/detail/1914>

4.2.2 Role

Advisory groups should be engaged at the planning and design stage of an evaluation synthesis wherever possible. They should continue their engagement throughout the synthesis process and be invited to provide feedback and review in explicitly defined stages. The roles of the advisory group include (these roles can vary):

Planning and design (if the advisory group is formed during the planning phase):

- Support the management group in determining synthesis objectives, thematic scope and guiding questions
- Reviewing the integrity of documents, such as Terms of References (TORs), inception reports and protocols
- Providing technical input and validating methodologies and approaches as described in the inception report and protocols

Conduct:

- Providing documentation (particularly documents not publicly available) during the search phase of the synthesis
- Reviewing the substance and relevance of initial or interim findings
- Reviewing draft deliverables, such as synthesis reports and briefs

Communication and uptake

- Supporting efforts to ensure the use and uptake of the synthesis and the final deliverables by key stakeholders in different forums and networks

4.3 What are the key considerations for engaging advisory groups?

- **Composition:** Advisory groups should be diverse and inclusive. There should be an appropriate gender, geographic, thematic and sectoral representation. They should be composed of individuals or institutional representatives reflecting the synthesis objectives and the thematic scope. This may include thought leaders or thematic experts. It may also include practitioners or policy makers. Furthermore, it may include representatives of stakeholder groups that form part of the audience targeted by the synthesis. Finally, consideration should be given to methodological experts.
- **Recruitment:** It is important to be explicit about the roles and responsibilities of potential advisory group members. Simple Terms of Reference can be useful in helping those invited to be part of an advisory group to have a sense of their responsibilities, time commitment, and estimated points of engagement.
- **Size:** There is no ideal size, and this may vary according to the availability of invited advisory group members and the size and scope of the synthesis. Consideration should be given to the balance between (1) ensuring a diversity of voices and perspectives (including representation by gender, geography, thematic backgrounds, organization type (civil society, government, etc.) and other

factors); and (2) ensuring manageability of the process, in terms of coordination and review/feedback.

- **Coordination and management:** There should be a clear focal point for engagement with the advisory group. Normally, this would be the synthesis team leader or managing focal point from the commissioning UN entity.

4.4 Can the composition of advisory groups vary for different synthesis types?

The above considerations apply to advisory groups for all three synthesis types. There are some additional considerations:

- **Single agency syntheses:** Consider thematic, geographical and practitioner expertise from different organisational levels and from different sections (e.g., Regional, Country Offices, and Headquarter representation).
- **Multi-agency syntheses:** All contributing UN agencies should have representation on the advisory group.
- **Syntheses of UN evaluations and other types of evidence:** It is even more important to include voices and perspectives external to the UN, including those with thematic, methodological, policy or practitioner expertise.

Useful resources

Bakrania, S. (2020). Methodological briefs on evidence synthesis. Brief 5: Commissioning and managing an evidence synthesis project, *Innocenti Research Briefs*, no. 2020-05, UNICEF Office of Research - Innocenti.

Better Evaluation/Global Evaluation Initiative (GEI). (n/a). *Advisory groups*.

<https://www.betterevaluation.org/methods-approaches/methods/advisory-group>

Lasserson, T.J., Thomas, J. Higgins, J.P.T. (2023). Section 1-3-2: Working with consumers and other stakeholders. In Higgins, JPT., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, MJ., Welch, VA. (eds.). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). Cochrane, 2023. Available from www.training.cochrane.org/handbook.

International Fund for Agricultural Development (IFAD). (2021). *The evaluation advisory panel of the Independent Office of Evaluation of IFAD*.

<https://ioe.ifad.org/documents/38714182/44720058/IOE+Evaluation+Advisory+Panel+-+booklet.pdf/f66e70d5-bb1f-24e6-5d7c-f675fc881a48?t=1645544053104>

Loud, M. L. (2018). *How do we use advisory groups effectively in evaluation?* Better Evaluation/Global Evaluation Initiative (GEI). <https://www.betterevaluation.org/blog/how-do-we-use-advisory-groups-effectively-evaluation>

United Nations Evaluation Group (UNEG). (2016). Standard 4.6: Stakeholder engagement and reference groups. *UNEG Norms and Standards*. <https://www.unevaluation.org/document/detail/1914>

Voegeli, C., Schmitt-Boshnick, M., & Krupa, E. (2021). Evaluation advisory groups: Considerations for design and management. *Canadian Journal of Program Evaluation/La Revue canadienne d'évaluation de programme*, 36.2 (Fall/Automne), 227-235. doi: 10.3138/cjpe.69949

World Food Programme (WFP). (n.d.). Technical note: evaluation products.
<https://docs.wfp.org/api/documents/WFP-0000130020/download/>

5. Developing a Terms of Reference for commissioning a synthesis

5.1. What is the purpose of a Terms of Reference?

A Terms of Reference (ToR) is necessary when commissioning a consultant, a group of consultants or an organization or firm to conduct the synthesis. The ToR provides an overview of what is expected in an evaluation synthesis and serves as the foundation for a contractual agreement between the UN commissioning entity and consultant/s, organization or firm. It establishes the background, purpose, objectives, initial synthesis questions, intended audience, proposed approach, timelines and deliverables, expected skills and qualifications needed in a synthesis team, and an indication of the expected level of effort for bidders.

The ToR should be developed after the initial conceptualization and preliminary scoping have been undertaken (see sections 4 and 5). This is because it is useful to have an initial indication of the existing evidence and syntheses (to avoid repetition), synthesis objectives, questions, scope and approach, even if these will be further developed and refined during the inception stage by the commissioned consultant/s, organization or firm (see section 5).

5.2. What should a Terms of Reference include?

ToR content and format may vary depending on organizational requirements and the type of assignment. However, the following components can be used as a template for developing a ToR for evaluation synthesis.

- **Background:** Briefly state the background and context of the proposed synthesis, what is already known in the topic of interest, the progress made so far (for SDG syntheses), and the gap and why the current synthesis is being done or implications for synthesis.
- **Purpose and objective of the synthesis:** State why and for whom the synthesis is being done. Include considerations on the target audience for the synthesis and the intended use.
- **Existing evidence base and syntheses:** Offer an up-to-date and concise summary of what is known on the topic, the existing evidence base and any gaps within current knowledge.
- **Guiding synthesis questions:** Guiding synthesis questions (preliminary questions) could be developed to inform the inception period. Provide as much definition as possible at this stage, and if necessary, explain that these can be refined during the inception phase in collaboration with the commissioned consultant/s, organization or firm.
- **Synthesis scope:** Based on preliminary scoping, define (as far as possible) the temporal, thematic and geographic scope, and inclusion and exclusion criteria for the types of evidence, and the size of the universe to be included in the synthesis
- **Approach and Methodology:** Propose (with as much definition as possible) the synthesis methods. These should align with the synthesis purpose and questions proposed, and the nature of evidence that is expected to be included. At the very least, it is useful to provide some guidelines on this, even if the commissioned consultant/s, organization or firm will fully refine the methodology during the inception phase. This section can include detail on the type of products

(narrative synthesis, evidence maps, systematic review), the development of a theory of change, search strategy, inclusion and exclusion criteria, abstract screening stage, full text screening stage, use of data extraction forms, data management and coding process, data analysis strategy, review process for draft products including the role of engagement committees. This section can outline the use and application of frameworks and mnemonics to guide the synthesis such as the Population, Intervention, Comparison, Outcome (PICO) approach or the use of a Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) checklist and diagram.

- **Structure of team and responsibilities:** Outline the skills and competences the synthesis requires including experience and delivery of appropriate forms of synthesis, sectoral expertise, language proficiency, database accessibility and experience, and demonstrable experience of delivering quality on time and to budget. The ToR should state the synthesis team's composition and minimal eligibility criteria. It may be useful to define the structure and level of effort anticipated for the synthesis, including the recommended size of the synthesis team, in alignment with the scope of the assignment.
- **Timeline and deliverables:** Describe the main phases of the synthesis process, including:
 - Phase 1: Inception phase – where the inception report, approach paper or protocol is developed with a fully defined scope, synthesis questions and methodology, possibly including inclusion and exclusion criteria, pilot coding, search steps, search terms,
 - Phase 2: Evidence collation phase – where systematic searches are conducted, initial evidence mapping, and the results generated are screened against the eligibility criteria.
 - Phase 3: Quality appraisal and data extraction phase – where included documents are quality appraised and where data is extracted from them
 - Phase 4: Analysis and synthesis phase – where the findings are synthesized and reported on, possibly including a final critical appraisal step.
 - Phase 5: Communication and uptake phase - where the synthesis output (a report or otherwise which will include recommendations and/or lessons learned) is communicated to key stakeholders.
- **Management and governance structure:** State the governance and accountability structures for the activity. It is useful to explain roles and responsibilities of the consultant/s, organization or firm and the management and quality assurance structure, which will include focal points from the commissioning UN entity and the advisory group (see section 6).
- **Duty station:** Explicitly state from where the work will be conducted and the balance between office and remote work (including perhaps the synthesis team's chosen location).
- **Payment schedule:** This section should outline the deliverable schedule, including the deliverable dates and percentage of funds to be transferred in each installment. The payments should align with the deliverables.
- **Application requirements for technical and financial proposals:** If applicable, the ToR should explain how the proposals will be evaluated. For example, proposals could be evaluated based on a combination of technical and financial considerations including the need to meet the mandatory criteria. The percentage score for each proposal should be stated (for example 70% for technical quality of the proposal vs 30% for the financial proposal).
- **Assessment of proposals:** If applicable, the ToR should state the assessment criteria for the technical proposals

- **Annexes:** Annexes could include supplementary information useful for bidders, such as preliminary scoping results, key strategic documents, draft data extraction forms,

Useful resources

Bakrania, S. (2020). Methodological Briefs on Evidence Synthesis: Brief 5 – Commissioning and managing an evidence synthesis project, Innocenti Research Brief 2020-03. Florence: United Nations Children’s Fund Office of Research – Innocenti. <https://www.unicef-irc.org/publications/1081-methodological-briefs-on-evidence-synthesis-brief-5-commissioning-and-managing-an.html>

UNIFEM: Guidance Note on Developing Terms of Reference (ToR) for Evaluations: https://www.endvawnow.org/uploads/browser/files/UNIFEM_guidance%20note_evaluation_ToR.pdf

Better Evaluation/Global Evaluation Initiative (GEI). [Terms of reference](#) | [Better Evaluation](#)

6. Defining evaluation synthesis questions

6.1. How can synthesis questions be developed?

Synthesizing evaluative evidence is most easily done when the component evaluations are focused on similar subjects and have comparable evaluation questions. When data has been collected, triangulated and analyzed to answer a common set of questions, the synthesis process is greatly simplified.⁹ Some agencies are purposeful in standardizing their evaluation questions across certain types of evaluations in order to facilitate syntheses. In some cases, the questions may be based on synthesizing evidence from evaluations on the OECD DAC Evaluation Criteria (relevance, coherence, effectiveness, efficiency, impact and sustainability).¹⁰ When this is not the case, then synthesis questions should be developed based on a clear conceptual framework developed following a scoping process, keeping in mind the purpose of the synthesis and intended use.

Consideration should be given to the evidence base and what questions can be answered given the existence of reliable, high quality evaluative evidence. Consideration should also be given to timelines and manageability. The scoping process (see section 7) will give some indication of the questions that can be answered based on the availability of appropriate evidence, the time and resources available to conduct the synthesis, and the availability of existing syntheses that answer the same or similar questions.

There should also be alignment between the synthesis questions and eligibility criteria (see section 9). Considerations on the sample, phenomenon of interest and the types of evidence to be included will help structure the synthesis questions.

As such, the process of defining questions may occur iteratively in line with the scoping stage.

- **As part of the initial conceptualization of a synthesis:** Some initial questions or broad objectives may be proposed, in line with standardized evaluation questions, or questions that respond to an organizational request or policy priority. These initial questions may help guide the initial scoping, which in turn may help further refine the synthesis questions.
- **Prior to developing a ToR for commissioning syntheses:** The broad objectives or more refined synthesis questions could be included in a ToR to help potential firms or consultants submit proposals that are aligned with expectations.
- **As part of the inception period of a synthesis:** During the inception period, synthesis teams may continue to refine objectives and questions as part of the inception scoping phase. The final synthesis questions should be defined in the inception report or protocol.

⁹ See WFP's synthesis of evidence from policy evaluations, for example.

¹⁰ Refer to OECD DAC guidance on Evaluation Criteria for more information:

<https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>

6.2. What kinds of questions can evaluation synthesis answer?

6.2.1 What works?

These questions typically explore the **effectiveness** and/or **impact** of interventions, activities, projects, programmes, strategies and policies. In other words, they look at what has worked well (and conversely, what has not worked well). Effectiveness is measured by the extent to which activities, projects and programmes achieved their stated objectives (which could be defined in terms of results, outputs or outcomes). Impact is measured by the change resulting from the intervention in the intended beneficiaries, or on norms, policies or systems. This is usually through rigorous impact evaluation methods based upon counterfactual analysis or other methods that assess contribution.¹¹

Box x: Examples of effectiveness and impact type questions in single and multi-agency syntheses

- What interventions and approaches to SDG5 have worked well and which have not? (From a multi-agency UN evaluation synthesis of SDG5)¹²
- Based on evaluative evidence, to what extent have United Nations agencies contributed to progress towards the eight SDG 6 targets? (From a multi-agency UN evaluation synthesis of SDG6)¹³
- What evidence of success is found by the evaluations regarding initiatives that tackle gender disparity and exclusion?? (From a multi-agency UN evaluation synthesis of SDG4.5)¹⁴

For **single agency or multi-agency syntheses** drawing upon UN evaluations, questions of effectiveness may relate to what has worked well or not in terms of evidence on explicit results, outputs or on outcomes. In some cases, UN evaluations may use rigorous techniques to measure the impact (through counterfactual design) or contribution (through rigorous contribution analysis) of interventions on intended beneficiaries, norms, policies or systems.

Syntheses of UN evaluations and other types of evidence will likely draw upon external research, such as impact evaluations, which use rigorous counterfactual design to assess intervention impact. Therefore, questions may be more explicit in their exploration or assessment of causal impact.

Box X: Example of ‘what works’ type questions in a synthesis of UN evaluations and other evidence

¹¹ Refer to OECD DAC guidance on Evaluation Criteria for more information on how effectiveness and impact are defined: <https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm>

¹² UNWomen. (forthcoming). Are We Getting There? A synthesis of UN system evaluations of SDG5. UNWomen, UNDP, UNICEF, UNFPA, WFP.

¹³ United Nations Children’s Fund. (2021). Evaluation Synthesis of United Nations System and Development Bank Work Towards SDG 6’. New York: UNICEF Evaluation Office.
https://evaluation.iom.int/sites/g/files/tmzbd1151/files/docs/resources/SDG%206%20Evaluation%20Synthesis%20Report_Final.pdf

¹⁴ Mundy K. & Proulx, K. (2019). Making evaluation work for the achievement of SDG 4 target 5: Equality and inclusion in education. Paris: UNESCO Internal Oversight Service Evaluation Office.
<https://unesdoc.unesco.org/ark:/48223/pf0000370558>

All these synthesis questions are taken from Global SDG Synthesis Coalition’s synthesis of evidence on the Partnership Pillar of the SDGs:¹⁵

- Which trade, finance, technology, systemic issues, and capacity building interventions are most effective in improving and accelerating SDG-17 indicators?
- What is the impact of trade, technology, finance, systemic issues and capacity building policies, programs, and interventions on SDG-17 indicators?
- How do gender and age considerations, in policies, programs, and interventions related to SDG-17 (e.g., gender mainstreaming, the sex and age of program participants, gender norms, women’s decision-making power and other gender and age considerations) influence positively or hinder the effects of trade, finance, technology, systemic issue, and capacity building interventions?

Consideration should be given to the principle of **Leave No One Behind**. Therefore, sub-questions could be incorporated to interrogate the impact or effectiveness of interventions on different target sub-groups considered to be the furthest left behind in any given context.

6.2.2 How or why something works?

These questions typically explore the factors (enabling or hindering) that determine progress towards objectives. These could relate to contextual factors that affect implementation or the achievement of objectives, or design and implementation issues. Questions related to interrogating issues of coherence, efficiency and sustainability could also be incorporated.

For **single agency or multi-agency syntheses**, how or why type questions could draw upon the qualitative analysis of project or programme implementation, process or performance commonly offered in UN evaluations.

¹⁵ de Hoop, T. et al., (2023), What works to accelerate progress on the Partnership Pillar of the Sustainable Development Goals: A synthesis of evaluative evidence. New York: The Global SDG Synthesis Coalition. <https://www.sdgsynthesiscoalition.org/sites/default/files/2023-09/Synthesis%20of%20the%20Partnership%20Pillar%20of%20the%20SDGs%20-%20THE%20FULL%20REPORT.pdf>

Box X: Examples of ‘how’ or ‘why’ type questions in single- and multi-agency syntheses

All these synthesis questions are taken from a multi-agency UN evaluation synthesis of SDG6¹⁶. This synthesis organized the questions in terms of the OECD DAC Evaluation Criteria.

- **Enabling or hindering factors:** What are the most common facilitating and hindering factors identified for progress towards SDG6?
- **Relevance:** How well-aligned were interventions with SDG 6 targets and with key commitments?
- **Relevance with an LNOB element:** How well were gender, equity and accountability to affected populations (AAP) integrated into interventions?
- **Coherence:** To what extent was there evidence of United Nations and sister agencies coordinating and linking efforts between the SDG 6 target areas (WASH, climate, water resource management)?
- **Sustainability:** To what extent are the evaluated interventions sustainable or likely to be sustained?

Syntheses of UN evaluations and other types of evidence may include external evidence in the form of qualitative research that explicitly analyses the implementation or performance of interventions, or on the evaluations of other bi- and multi-lateral development agencies.

Box X: Example of ‘how’ or ‘why’ type questions in a synthesis of UN evaluations and other evidence

All these synthesis questions are taken from Global SDG Synthesis Coalition’s synthesis of evidence on the Partnership Pillar of the SDGs.¹⁷

- Why and how are some interventions more or less successful in achieving progress towards SDG-17?
- What does the evidence say about what does and does not work to implement effective trade, finance, technology, systemic issue, and capacity building policies, programs, and interventions under SDG-17?
- Under what conditions (e.g., income status, human development status, region, gender, age) were the interventions most effective?

¹⁶ United Nations Children’s Fund. (2021). Evaluation Synthesis of United Nations System and Development Bank Work Towards SDG 6’. New York: UNICEF Evaluation Office.
https://evaluation.iom.int/sites/g/files/tmzbd1151/files/docs/resources/SDG%206%20Evaluation%20Synthesis%20Report_Final.pdf

¹⁷ de Hoop, T. et al., (2023), What works to accelerate progress on the Partnership Pillar of the Sustainable Development Goals: A synthesis of evaluative evidence. New York: The Global SDG Synthesis Coalition.
<https://www.sdg-synthesis-coalition.org/sites/default/files/2023-09/Synthesis%20of%20the%20Partnership%20Pillar%20of%20the%20SDGs%20-%20THE%20FULL%20REPORT.pdf>

As before, consideration should be given to the principle of **Leave No One Behind**. Therefore, questions should be designed to interrogate:

- The incorporation of LNOB (or equity or equality) principles in the design or implementation of interventions (e.g. the explicit inclusion of certain groups, how resources are allocated between different groups, how design aspects respond to the needs of different groups, whether the intervention is compatible with inclusion and equity norms).
- The extent to which contextual factors (e.g. in hard-to-reach areas or fragile contexts) affect implementation or the achievement of objectives.

7. Mapping or scoping the evaluative evidence base

7.1. What is the purpose of mapping or scoping the evaluative evidence base?

When developing or commissioning an evaluation synthesis, some initial thinking and research may be necessary to help define the synthesis objectives, questions and thematic scope. An initial mapping or scoping can indicate how much potential evidence is available on a particular topic, theme or intervention. If this initial mapping or scoping suggests a potentially large universe of evaluations, the scope may be further refined to ensure manageability. Conversely, the scoping or mapping may indicate a small body of evidence that is not sufficient to answer the stated objectives or research questions, in which case decisions can be made on expanding the scope, or not pursuing the synthesis altogether.

7.2. When in the synthesis process can a scoping occur?

There are different points at which it may be useful to undertake a scoping. Indeed, scoping may not be a one-off exercise and could progress in an iterative fashion through the early stages of a synthesis project. In all cases, scoping is likely to be a collaborative process, with feedback loops built in to ensure that the final synthesis meets requirements and that there is final agreement on the specific objectives, questions and scope.

- **As part of the initial conceptualization of a synthesis:** Scoping could help conceptualize the synthesis at very initial stages when the precise objectives and thematic scope are still being discussed. The scoping findings could be presented in the form of an initial concept or scoping note, which can then form the basis for consulting on or proposing the objectives, questions and thematic scope in response to requests by the governing bodies of evaluation offices or senior management.
- **Prior to developing a ToR for commissioning syntheses:** When the broad objectives or general thematic focus of the synthesis is known, a scoping could help provide some further definition for inclusion in a ToR. This will help potential firms or consultants submit proposals that are more realistic.
- **As part of the inception period of a synthesis:** It is standard practice for synthesis teams to undertake a scoping as part of the inception period of a synthesis. This helps further refine objectives, questions, eligibility criteria and search strategies for inclusion in the inception report or protocol.
- **As part of a formal scoping review:** Formal scoping reviews can be specific deliverables. A formal scoping review may be undertaken as an initial and substantive phase of a synthesis activity. A scoping report providing a detailed analysis of the quantity and nature of evaluative evidence can be delivered, and then decisions can be taken on how best to proceed with the actual synthesis as a second phase.

7.3. How to undertake a mapping or scoping

Depending on the time and resources available, scoping exercises may range from ad hoc and informal, to more formal scoping reviews, or somewhere in between.¹⁸

Ad hoc approaches entail quick and informal searches in evaluation databases and repositories of other types of evidence. The searches used are based on initial concepts, themes, sectors or interventions envisaged.

- For **single agency syntheses**, this will involve searches in the agency's evaluation database. These could be keyword searches or searches using existing thematic or non-thematic filters, depending on the functionality of the database.
- For **multi-agency syntheses**, these searches could be undertaken in the UNEG Database of Evaluation Reports, or in a selection of contributing agency databases.
- For **syntheses of evaluations and other types of evidence**, a selection of UN evaluation databases and key external databases could be used. These searches will give an indication of the quantity of evidence available. Some UN evaluation databases provide thematic and non-thematic metadata, which helps provide an initial and useful indication of the evaluation types and other attributes (e.g., thematic coverage, relevant SDG etc.). Where this metadata is not readily available, a sample of the search results could be reviewed to ascertain the attributes. Where other types of evidence are included in the synthesis, searches will need to be conducted in non-UN repositories. A sample of repositories could be chosen, e.g. an academic database, an institutional repository, a grey-literature database.

More formal scoping reviews draw on systematic and comprehensive approaches to searching and exploring the extent of evaluative evidence for a topic, theme or sector. In comparison to ad hoc approaches described above, the searches may incorporate a more formal and comprehensive search strategy based upon a pre-defined or pre-existing conceptual framework. All evaluations and studies would be coded for their attributes (not findings) and the analysis would be provided in substantive scoping report with recommendations of where a subsequent synthesis can usefully focus. An evidence gap map, providing a visual representation of the type and distribution of evidence, can be produced as part of this formal scoping review.

In all cases, it is useful to map and **consider existing syntheses** in any decisions on the objectives and scope. This is to minimize duplication with existing work. It may be the case that syntheses already exist that fully or partly respond to the objectives, thematic scope or questions envisaged for a new synthesis. An assessment of duplication or overlap can be considered in decisions on defining the scope. Searches for UN evaluation syntheses can be conducted in the UNEG Database of Evaluation Reports or individual UN agency evaluation databases. Syntheses of other types of evidence can be found in external repositories. Suggestions for synthesis repositories are provided below.

¹⁸ Bakrania, S. (2020). *Methodological Briefs on Evidence Synthesis: Brief 3 – Developing and designing an evidence synthesis product*, Innocenti Research Brief 2020-03. Florence: United Nations Children's Fund Office of Research – Innocenti, Florence.

Useful resources

General guidance on scoping

Armstrong, R. et al., (2011). “Scoping the Scope” of a Cochrane Review, Journal of Public Health, 33, 1, pp. 147–150. <https://doi.org/10.1093/pubmed/fdr015>

Scoping reviews: what they are and how you can do them? A series of training videos published by the Cochrane Collaboration. <https://doi.org/10.1093/pubmed/fdr015>

Evaluation databases

UN Database of Evaluation Reports: <https://www.uneval.org/document/library>

DAC Evaluation Resource Centre: <https://www.oecd.org/derec/home/?hf=5&b=0&s=score>

Synthesis Databases

Social Systems Evidence. Contains 1000s of systematic reviews categorized by SDG:

<https://www.socialsystemsevidence.org/>

Campbell Systematic Reviews Journal. An open access and searchable database of social science systematic reviews and evidence gap maps published after peer review by the Campbell Collaboration:

<https://onlinelibrary.wiley.com/journal/18911803>

3ie’s Development Evidence Portal: An open access and searchable database of impact evaluations, systematic reviews and evidence gap maps published and/or compiled by 3ie. All are searchable by SDGs: <https://developmentevidence.3ieimpact.org/>

8. Protocol/inception reports

8.1 What is the purpose of an inception report or protocol?

The inception report or protocol is vital to the development and review process for an evaluation synthesis product. It explicitly outlines the scope, synthesis questions, methodology and timelines, and ensures that all the important decisions about the synthesis process have been made in advance.

8.2 Should inception reports or protocols be published?

In the broader evidence synthesis community, protocols are registered with evidence synthesis repositories or are published on organizational websites. In some cases, registration with a repository includes a peer review element. The intention is to ensure transparency and reproducibility, to reduce bias, and to ensure that other synthesis teams do not duplicate efforts.

This might not be appropriate for **single or multi-agency evaluation syntheses** that only include UN evaluations. However, it is recommended that UNEG member agencies at least publish inception reports and protocols on their organizational repositories and communicate these to the broader synthesis community through social channels and Communities of Practice.

For **syntheses of UN evaluations and other types of evidence**, it may be more appropriate to register the protocol in an established repository. This is especially important if the synthesis includes external academic research and grey literature and is being undertaken as a global public good.

8.3 What should be included in an inception report or protocol?

The inception report or protocol will usually include the following information:

- **Title:** The title should be informative and clearly indicate the topic of the evaluation synthesis. Titles should not be phrased as questions; there should be congruency between the title, synthesis purpose, objectives/questions, and eligibility criteria. It is useful to structure the title of the protocol (and resulting synthesis) according to the eligibility criteria (see section 9, which proposes the SPIDER eligibility framework).
- **Background:** The background should describe and situate the phenomena of interest under review and the context of the synthesis. The background should also explain the justification and rationale for conducting the evaluation synthesis. If there is an existing evidence synthesis on the topic, it should specify how the proposed synthesis will differ.
- **The synthesis objectives and questions:** Clarity in the synthesis questions assists in developing a protocol and, ultimately, the conduct of the synthesis properly. The review question/s guides and directs the development of the specific review criteria, facilitates more effective searching, and

provides a structure for the development of the full synthesis (see section 4 for more guidance on developing synthesis questions).

- **A conceptual framework or theory of change:** setting out how the intervention, initiative, programme of project works, for whom and under what circumstances
- **Eligibility criteria:-** This section of the protocol details the basis on which studies will be considered for inclusion into the proposed synthesis and should be as clear and unambiguous as possible (see section 9, which proposes the SPIDER eligibility framework).
- **Methods:** This section details the methods, including the search strategy, the databases or sources to be searched, the purposive sampling strategy (if appropriate), the process for screening documents, the process for quality appraising evaluation evaluations and studies (and their associated appraisal tools); the process for extracting data, including the types of data to be extracted; and the methods to be used for analyzing or synthesizing studies. The research team could use or adapt the conceptual framework indicated in the ToR to support the synthesis process and should explicitly state this in their protocol or inception report. Any software or tools that will be used for extraction, analysis or synthesis should be explained in the protocol.
- **Proposed timeline:** include a detailed workplan which sets out the agreed timeline and deliverables which may align to the timeline indicated in the TOR.
- **Appendices** may include items such as a screening checklist, a list of databases searched and the search engine, a list of the types of data to be extracted, and the quality appraisal criteria.

Useful resources

Bakrania, S. (2020). *Methodological Briefs on Evidence Synthesis: Brief 3 – Developing and designing an evidence synthesis product*, *Innocenti Research Brief 2020-03*. Florence: United Nations Children’s Fund Office of Research – Innocenti. <https://www.unicef-irc.org/publications/pdf/IRB%202020-03.pdf>

Aromataris E, Munn Z (Editors). *JBIM Manual for Evidence Synthesis*. JBI, 2020. Available from <https://synthesismanual.jbi.global>. <https://doi.org/10.46658/JBIMES-20-01>

9. Eligibility criteria

9.1 What is the purpose of eligibility criteria?

The eligibility criteria for the synthesis transparently defines what kinds of evidence are included and excluded from the synthesis. Rather than leaving this until the inception phase of a synthesis, it is useful to begin developing thinking on this at the scoping/mapping phase, and then to continue refining throughout the synthesis process in an inductive/deductive manner. The scoping and mapping phase provides an opportunity to refine the synthesis questions and thematic scope. At the end of this process, the synthesis objectives and questions should outline the problem to address, specify the population and context to which the questions apply, and identify the themes, topics or interventions of interest.^{19,20} Moving further, the inception report or protocol should define the eligibility criteria for the synthesis, which determines what kind of evidence is included and excluded. The eligibility criteria provide a useful framework that can help define the synthesis questions (see section 4) and the search strategy (see section 10), and all these aspects should be aligned.

9.2 What frameworks are available for developing eligibility criteria?

Synthesis exercises that are conducted for accountability purposes may be based on common types of evaluations (e.g. country programme evaluations) that have been designed with standardized evaluation questions to facilitate synthesis, among other things. In this case, there may be less creativity in defining synthesis questions as they will be based on the standardized evaluation questions.

The broader the scope of the synthesis and size of evidence base, the more thorough and detailed the definition of eligibility criteria will need to be.

The SPIDER framework is a useful approach for structuring evaluation synthesis questions for, formulating the eligibility criteria, and for developing the search strategy. This framework is particularly suited to qualitative and mixed-method syntheses.²¹

¹⁹ Armstrong, R., et al. (2011). "Scoping the Scope" of a Cochrane Review, *Journal of Public Health*, 33, 1, 2011, pp. 147–150. <https://doi.org/10.1093/pubmed/fdr015>

²⁰ Bakrania, S. (2020). *Methodological Briefs on Evidence Synthesis: Brief 3 – Developing and designing an evidence synthesis product*, Innocenti Research Brief 2020-03. Florence: United Nations Children's Fund Office of Research – Innocenti. <https://www.unicef-irc.org/publications/pdf/IRB%202020-03.pdf>

²¹ Cooke A, Smith D, & Booth A. (2012) Beyond PICO: The SPIDER Tool for Qualitative Evidence Synthesis. *Qual Health Res.*, 22, 10, pp.1435–1443. <https://doi.org/10.1177/104973231245293>

SPIDER	Description
S - Sample	<p>Which groups of people, population, contexts or settings are of interest? The synthesis may focus on specific groups or maintain a broader scope to look at regions or countries of interest.</p> <p>A factor for evaluation synthesis is to consider from which UN or other implementation agencies are evaluations being considered for inclusion.</p>
PI – Phenomenon of Interest	There may be certain themes, interventions, activities, programmes, policies, or behaviors or experiences of interest.
D – Design	<p>What evaluation types or study design are of interest for inclusion in the synthesis?</p> <p>For evaluation types, it would be useful to specify whether the inclusion of all or specific types are appropriate (e.g. project, programme, thematic, country portfolio, impact, institutional effectiveness or strategy evaluations).</p> <p>For other evidence types, it would be useful to specify the research design or data collection methods of interest (e.g. randomized controlled trial, quasi-experimental, observational, survey, case study).</p> <p>The publication period of evaluations or other evidence types to be included should be considered.</p> <p>A further factor to consider is whether only evaluations of a certain quality rating or above are being considered for inclusion.</p> <p>Language may also be another factor to consider. Is eligibility restricted to evaluations or studies of certain languages, or will no language restrictions be applied?</p>
E - Evaluation	What is being measured in evaluations and studies that is of interest to the synthesis questions? These could be measures of success or progress towards objectives, measures of programme or project effectiveness or impact used in evaluations, or other measures of other evaluation criteria (relevance, coherence, efficiency and sustainability). In some cases, this could refer to measured experience or perceptions from participating in an intervention.
R – Research type	This could be quantitative, qualitative or mixed method.

9.3 Examples of eligibility criteria

Below are examples of how UN evaluation syntheses have developed eligibility criteria based upon the SPIDER framework.

Box x: SPIDER eligibility criteria in a single agency synthesis²²

The example below is from an UNDP IEO synthesis on interventions supporting job creation in the micro, small and medium enterprise sectors in the Arab States. It includes both UNDP evaluations and external impact evaluation evidence.

	Inclusion criteria	Exclusion criteria
Publication dates	UNDP evaluations: 2018-2024 External sources from impact evaluations and systematic reviews: 2018-2024	
Publication accessibility	Publicly available or shared with the synthesis team Published in English, Arabic and French	
Sample (population)	MSMEs (including startups) in Algeria, Bahrain, Djibouti, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Palestine, Saudi Arabia, Somalia, Sudan, Syria, Tunisia, and Yemen Women and youth	Cooperatives and social enterprises MSMEs focused on Oman, Qatar, and United Arab Emirates
Phenomena of interest (interventions)	The following type of MSME support interventions and approaches: <ul style="list-style-type: none"> • Policy and regulatory reforms • Access to markets • Infrastructure development • Financial support or access to the market • Promoting innovation and entrepreneurship • Networking and partnerships, • Digitalization and technological support • Capacity building and skills development • Women empowerment • Social inclusion • Gender equality • Environmental sustainability 	Evaluations or studies not related to MSME support interventions
Design (evaluation types)	Types of evaluations or study designs: <ul style="list-style-type: none"> • Independent Country Programme Evaluations (ICPEs) • Thematic and corporate evaluations • Decentralized project/program/outcome evaluations • Regional evaluation • Multi-country evaluations, • Strategy/policy evaluations, • Impact evaluations and other evaluation types, including case studies. • Impact evaluations from experimental and non-experimental (quasi-experimental) studies; and 	Midterm evaluations Non-evaluative documents (targeted literature reviews, programmatic or policy papers, conference proceedings, blogs)

²² UNDP IEO (forthcoming). Supporting job creation in the micro, small and medium enterprise sectors in the Arab States: An evaluation synthesis for UNDP's Regional Bureau of Arab States.

	<ul style="list-style-type: none"> • Systematic reviews and other rigorous syntheses, which may include rapid evidence assessments and rapid reviews <p>Publication dates:</p> <ul style="list-style-type: none"> • UNDP evaluations: 2018-2024 • External sources from impact evaluations and systematic reviews: 2018-2024 <p>Publicly available or shared with the synthesis team</p> <p>Published in English, Arabic and French</p>	
Evaluation (outcomes)	<p>Primary outcomes:</p> <ul style="list-style-type: none"> • Job creation • Income generation • Creation of decent jobs • Competitiveness (in terms of propensity to export, profitability, and investment) <p>Intermediate objectives/outcomes</p> <p>Macro-level policy or legislative change, partnerships and market access, access to finance, adoption of digital and new technologies, capacity and skills</p>	Evaluations or studies not reporting outcomes relating to MSME support interventions
Research type	<p>Evaluations or studies conducted using:</p> <ul style="list-style-type: none"> • Quantitative • Qualitative • Mixed methods research 	

Box x: SPIDER eligibility criteria in a multi-agency synthesis²³

The example below is from a Global SDG Synthesis Coalition synthesis related to the Peace Pillar of the SDGs, on the impacts of violence prevention programmes on homicides and conflict-related deaths and their implementation. It includes UN evaluations and external evaluative evidence.

Domain	Inclusion criteria
Publication dates	2019–2024
Publication accessibility	Published in English, Spanish, or French Publicly available or shared with the synthesis team
Sample	Focuses on sample(s) in low- and middle-income countries. We will also include studies which link interventions in high-income countries and outcomes in low- and middle-income countries.
Phenomenon of Interest	Assesses the impacts or implementation process of programmes pertaining to social inclusion, peace processes, and safe environments using impact, performance, and process evaluations
Design	Impact evaluations (randomized controlled trials/experiments or quasi-experimental studies with a comparison group), small-n impact evaluations (process tracing, contribution analysis, most significance change, outcome mapping, etc.), and performance or process evaluations (e.g., process evaluations or implementation science). Studies can be included if they cover both impact and performance/process elements (e.g., mixed-methods studies). Publication dates: 2019–2024 Languages: Published in English, Spanish, or French Publication accessibility: Publicly available or shared with the synthesis team
Evaluation	Outcomes related to violence prevention and peacebuilding at micro and macro levels. Gender considerations: Estimation of heterogeneous impacts by gender and examine gender considerations in performance and process evaluations
Research	Quantitative impact, and qualitative or mixed-methods performance and process evaluations focused on outcomes related to SDG indicators 16.1 and 16.4

Useful resources

Cooke A, Smith D, & Booth A. (2012) Beyond PICO: The SPIDER Tool for Qualitative Evidence Synthesis. *Qual Health Res.*, 22, 10, pp.1435–1443. <https://doi.org/10.1177/104973231245293>

²³ De Hoop et al. (forthcoming). The impacts of violence prevention programmes on homicides and conflict-related deaths and their implementation. New York: The Global SDG Synthesis Coalition.

10. Systematic searching

10.1 What is the purpose of systematic searching?

A systematic search is a comprehensive and transparent process to identify all evaluations and studies relevant to answering synthesis questions, minimizing the risk of missing critical information. The approach for systematic searching should be based upon a search strategy that is clear and documented to allow others to replicate (e.g., for future updates as part of a living synthesis). The results of the search should be well documented. Such transparency is critical to allow users to understand how evidence was identified, selected and analysis. Furthermore, systematic searching allows to minimize selection bias by following pre-defined and objective criteria to identify and include evidence. This helps avoid cherry-picking evaluations that might skew findings and ensure a balanced representation of available evidence.

10.1. How to develop a search strategy?

The search strategy is a key component of the inception report or protocol. It consists of:

- A statement of the databases within which the searches will be conducted
- The method for searching in databases, which may entail the use of key search, combinations of key search terms (known as search strings) and the use of database filters. However, the filters used in some databases (such as tags for thematic areas or SDGs) may not always be accurate or consistent within a single database or across multiple databases. Researchers should be mindful of this limitation and manually verify the accuracy of these filters if they choose to rely on them.

10.1.1 Databases for single or multi agency syntheses of UN evaluations

The primary sources of information for single-or multi-agency syntheses of UN evaluations are the UN agency organizational evaluation databases. The UNEG database of evaluation reports is an additional database. However, not all evaluations from individual agency databases are uploaded to the UNEG database. Therefore, researchers should verify with the UN agencies included in the sample to ensure they have obtained all evaluations published by each agency during the research period. Annex 1 includes a list of UN agencies and links to their evaluation databases.

10.1.2 Suggested databases for syntheses that include both evaluations and other types of evidence

In addition to the above-listed sources of information (UN evaluation databases) for the single agency or multi-agencies, it is important to locate the sources of information for other types of evidence (if the evaluation synthesis aims for this). Annex 2 provides some initial suggestions for repositories of external research, although this is not exhaustive. Many academic journal databases require subscription, although some articles are open access.

10.1.3 Developing a search strategy for internal UN databases and (where appropriate) for external research repositories

The starting point for developing a search method is to consider the synthesis questions, the conceptual framework and the eligibility criteria. For example, key terms included in the eligibility criteria (see section 9) relating to each of the components of the SPIDER framework can be used as the basis to develop search terms or search term combinations.

SPIDER framework component	Types of search terms
----------------------------	-----------------------

S – Sample	Search terms relating to the groups of people, populations, contexts or settings are of interest, alongside the period under analysis.
PI – Phenomenon of Interest	Search terms relating to the themes, development initiatives, interventions, behaviors or experiences of interest.
D – Design	Search terms defining the evaluation types or study design that are of interest
E – Evaluation	Search terms that further define the objectives or outcomes of interest
R – Research type	In some cases, it may be useful to further specify whether you are interested in qualitative, quantitative or mixed methods studies.

Search terms in each of these components can form a ‘block’ of search combinations, where the individual terms are combined using ‘AND’ or ‘OR’ Boolean combiners. Searches should also be conducted in multiple languages commonly used in UN evaluation reports, with a particular focus on French and Spanish alongside English.

Box x: Example search term combinations from a synthesis of UN evaluations and other types of evidence

The following is a sub-set of the search strategy from the Global SDG Synthesis Coalition’s Synthesis of evidence on the Partnership Pillar of the SDGs. Only a portion of the search strategy is shown here, relating to the ‘Phenomenon of Interest’ and ‘Design’ components of the framework. In this case, the Phenomenon of Interest was interventions relating to SDG17 (*trade, technology, finance, capacity building and systemic issues). These search term combinations were applied in academic and institutional databases external to the UN to search for impact evaluations, so the design terms relate to impact evaluation designs.

Phenomenon of Interest/Intervention: tax, OR trade, OR export, OR tariff, OR technolog*, OR digital technolog*, OR internet, OR phone, OR mHealth, OR mobile health, OR mobile money, OR mobile banking, OR finance, OR savings, OR self-help groups, OR market-led, OR market-based reforms, OR corporate social responsibility, OR certification schemes, OR organic, OR blockchain, OR laptop, OR business support services, OR public-private partnerships

Design: evaluation, OR impact evaluation, OR random* controlled trial, OR experiment, OR quasi-experiment, OR regression discontinuity, OR difference-indifferences, OR assessment, OR propensity score, OR systematic review, OR rapid review, OR evidence synthesis

In practical terms, there may need to be some flexibility in how the search strategy is applied. Individual databases have different levels of functionality. For example, bibliographic databases that hold academic journal articles often have complex and advanced search functionality, which allows complex search term combinations to be entered to specify the search. Search terms in each of the SPIDER framework

components can form a ‘block’ of search combinations, where the individual terms are combined using ‘Boolean combiners (AND, OR, NOT).

In contrast, many UN agencies and multi-agencies databases have limited functionality, which will require the search strategy to be adapted. Synthesis team members should explore the agency’s database, identifying its features and testing different search options. The resulting search strategy may be an iterative process, using the ‘best fit’ use of thematic search terms and available filters that work best on each database. From this, a search approach specific to that database could be developed and recorded.

The searches done in each database should be recorded (copied and pasted exactly as run) and annexed to the report. The number of search results should also be recorded and could be presented in the form of PRISMA diagram or any other preferred format in the final report.

10.2. What tools can be used to manage references?

Several reference management tools exist to help record and manage references. This is especially useful when dealing with many search results and evaluations or studies. These tools can also help identify duplicates (e.g. from searches in different databases that may generate duplicate results). At present, these work better for managing research studies. Common examples include Zotero (<https://www.zotero.org/>) and Endnote (<https://endnote.com/>).

Useful resources

Bakrania, S. (2020). Methodological Briefs on Evidence Synthesis: Brief 3 – Developing and designing an evidence synthesis product, Innocenti Research Brief 2020-03. Florence: United Nations Children’s Fund Office of Research – Innocenti. <https://www.unicef-irc.org/publications/pdf/IRB%202020-03.pdf>

Bakrania, S. (2020). Methodological Briefs on Evidence Synthesis: Brief 4 – Collating and analyzing studies for synthesis, Innocenti Research Brief 2020-03. Florence: United Nations Children’s Fund Office of Research – Innocenti. <https://www.unicef-irc.org/publications/pdf/IRB%202020-04.pdf>

Aromataris E, Munn Z (Editors). JBI Manual for Evidence Synthesis. JBI, 2020. Available from <https://synthesismanual.jbi.global>. <https://doi.org/10.46658/JBIMES-20-01>

Guidance on recording search results in a PRISMA diagram

Page, M.J. et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, 71. <http://www.prisma-statement.org/PRISMAStatement/FlowDiagram>

11. Screening

11.1. What is the purpose of screening

After conducting the systematic searches, the results need to be screened against the eligibility criteria to determine which evaluations and studies are included in the synthesis. This is a labour-intensive part of the process and is done by comparing each evaluation or study to the eligibility criteria and determining whether to include or exclude.

11.2. How to conduct a screening

Firstly, it is useful to have a screening tool to guide decisions on inclusion or exclusion. This tool will align with the eligibility framework. The screening itself is usually conducted in two phases: by title, thematic label and abstract/executive summary; and then by full text. At each stage, the number of evaluations or studies to screen reduces, but the level of detail involved in screening increases. This helps to reduce bias and improves consistency in the process.

11.2.1. Initial screening by title, database filter, executive summary or abstract

The purpose of the initial screening step is to quickly exclude evaluations or studies that fall outside of the eligibility criteria. This initial screening may be based on criteria that are easy to identify – e.g. thematic focus, regional focus, evaluation type, publishing date etc. The title of an evaluation, or its executive summary may not always provide this information explicitly. For UN evaluations, it may be sufficient to combine the searching and screening phases (see section 10) to produce an initial universe of evaluations for further full-text screening. For example, using key search terms or filters relevant to the eligibility criteria to narrow down the number of possible evaluations to be considered for inclusion. This may be sufficient when the synthesis objectives and questions are broad (e.g. multi-agency synthesis of progress towards SDG targets). If the objectives and synthesis questions are very specific (e.g., focused on very specific intervention types), it may be useful to screen further by looking at the executive summaries.

For syntheses that include other types of evidence, including external research evidence, it is the norm to screen by title and abstract and keywords, as relevant. The titles and abstracts of research studies are often very explicit on their thematic focus, types of interventions, populations targets, research design etc.

Any screening that progresses beyond searches and the reading of titles and abstracts should be conducted by at least two independent reviewers. They should screen at least a portion of the same evaluations and studies (known as double-screening), compare their decisions, reconcile any differences, and come to agreements about their decisions going forward.

11.2.2. Further screening by full text

The purpose of this second step is to look at the detail of the evaluation or study and to ensure that it meets the full eligibility criteria (i.e., that it is relevant to answering the synthesis questions). This screening may be on criteria that are more difficult to identify from the title, abstract or executive

summaries, e.g. specific intervention focus, outcomes, specific populations, specific regions or countries, specific research design used. For some evaluations or studies, it may be sufficient to undertake a strategic reading of the document, using document search functions to focus on specific sections of the document (e.g. those sections that speak to the methodology, the interventions, the populations targeted etc.). In some cases, when there is a rapid deadline, the full text screening and data extraction phases could be combined.

Full text screening should certainly be conducted by at least two independent reviewers. They should screen at least a portion of the same evaluations and studies (known as double-screening), compare their decisions, reconcile any differences, and come to agreements about their decisions going forward. Ideally, a third reviewer may be considered for special cases in which the two main reviewers cannot agree or when the discrepancy between their scores is significant.

11.2.3. What if the universe of evaluations is too large after searching or screening?

If searches in databases and any subsequent screening by title, database filter, abstract or executive summary still yields many evaluations, purposive sampling may be considered for qualitative evaluations or studies (see section 12).

11.3. Tools available for screening

Screening can be undertaken manually, either using Excel sheets to keep a record of screening criteria, decisions to include or exclude studies, and corresponding reasons for exclusion.

There are also several systematic review software tools that provide a platform for screening. These are particularly useful for allocating evaluations or studies to different members of a review team and for keeping track of which evaluations and studies are progressing through the screening stages. Examples of such tools include (but are not limited to) EPPI-Reviewer, Covidence, and SWIFT-Review.

More interestingly, many of these tools also include machine learning/artificial intelligence-based ‘priority screening’ assistants that help significantly reduce the screening workload. It must be noted that at present, the algorithms of these tools work on a reading of the title and abstracts of research studies, and therefore have not been sufficiently tested for UN evaluations. These work by recording screening decisions made on a sample of studies, and then ‘pushing’ more relevant studies to the top of the screening list. The review team can then screen to exhaustion (e.g. when studies are no longer being included) or decide to only screen a portion of the search results based on probabilities of inclusion (e.g. screening only those studies that have 60 percent probability or above of being included).

No such equivalent tool exists for UN evaluations yet. UNDP’s Artificial Intelligence for Development Analytics tool can help reduce the screening workload but only works for UNDP evaluations at present. It does this by identifying very relevant evaluations (or paragraphs from these evaluations) for corresponding search strategies. The results it generates (i.e. the paragraphs) can be used for screening or for forming the basis of a purposive sampling approach.

12. Purposive sampling for qualitative evaluation synthesis

12.1. What is the purpose of purposive sampling in a synthesis?

Sampling can be considered for qualitative synthesis of UN evaluations or external qualitative research when the initial universe of evaluations or research is too large or unmanageable for the time and resources available.

Following the principles of synthesis and existing global synthesis standards, purposive sampling should only be considered for qualitative synthesis. There are several justifications for this:¹

- UN evaluations and qualitative research studies include large amounts of rich and complex qualitative data. It can be overwhelming when dealing with large quantities of qualitative evaluations or research studies, which can reduce the depth of analysis.
- Sampling can ensure that thematic, geographic and other types of variation important to the synthesis questions are incorporated into the synthesis.
- Quantitative synthesis aims to be exhaustive (i.e., to include all evidence relevant to a synthesis question) to achieve statistical generalizability. This is not the case in qualitative synthesis, which aims for data saturation.

12.2. When does purposive sampling occur in the synthesis process?

Sampling will only be possible after the searching and initial screening stages, once the evaluations and studies resulting from searches have been initially screened at title, abstract or executive summary against the eligibility criteria (see section 11). Sampling can also only be undertaken if data on important characteristics or attributes of these evaluations and research studies are available. When this data is not readily available, it may be necessary to undertake an initial coding exercise before sampling.

12.2.1 When data on sampling criteria are readily available

UN evaluation databases often have meta-data that readily provides the characteristics or attributes of an evaluation (e.g., on regional/country coverage; on evaluation type; on sector/SDG/thematic focus). Many academic or institutional repositories of research studies also readily provide meta-data (e.g., subject or topic labels; on study type; on country focus). Where this is readily available, and where this can easily be exported (e.g., into an Excel file), sampling could feasibly be conducted without having to undertake an initial coding exercise.

12.2.2 When data on sampling criteria are not readily available

If data on evaluation and study characteristics or attributes are not readily available, then the following steps should be taken:

- Screen all search outputs against the inclusion criteria. If your searches result in a large number of outputs, then you may consider screening quickly against some of the key and easily identifiable inclusion criteria.
- Undertake a simple coding of the resulting studies, on characteristics that form the basis of the sampling criteria (e.g., evaluation or study type, geographic coverage,

sectors/interventions/topics addressed). (Use of the SPIDER framework can be applied at this stage).

12.3. What are the key considerations?

- What is the 'optimum' size of the universe of evaluations and/or research studies to be synthesized, considering the time and resources available? It can be useful to work backwards from this optimum figure and design the sampling approach accordingly.
- What are the key criteria to guide the purposive sample to ensure that the universe of evaluations and/or research includes the correct variations needed to answer the synthesis questions?
 - Are there certain evaluation types that need to be included in the sample?
 - Are there certain countries or regions that need to be included in the sample?
 - Are there certain sectors, topics or intervention types that need to be included in the sample?
 - Are there other important characteristics that should form the basis of the sampling criteria (e.g., population targeted)

12.4. What are the different approaches to purposive sampling?

Type of sampling	Description
Criterion sampling	<ul style="list-style-type: none"> ○ Studies are sampled based upon a pre-determined criteria (e.g., geography, evaluation or study type, sector/intervention/topic) ○ Studies are then analyzed according to the synthesis questions.
Stratified purposeful sampling	<ul style="list-style-type: none"> ○ This follows from criterion sampling, where each criteria becomes a sample. ○ As such, stratified samples are samples within samples. ○ It can be useful where the universe of evaluations is large, even after criteria sampling, and for focusing in on possible variations in topics of interest
Purposive random sampling	<ul style="list-style-type: none"> ○ Randomly select from the included studies ○ Random selection can be based on a random internet-based number generator, where the number corresponds to a numbered list of evaluations or studies. ○ Provides an unbiased way of selecting evaluations or studies, but there is less control over the variation of evaluations or studies included.
Snowball sampling	<ul style="list-style-type: none"> ○ Snowball sampling is more appropriate for research studies than evaluations. ○ It should not be used in isolation – it should be combined with a systematic search and another sampling type listed here. ○ It relies on locating seminal studies (e.g., through talking with experts or identifying a study that is regularly cited) ○ Forwards snowball sampling entails using the reference list/bibliography of studies to identify further relevant studies. ○ Backwards snowball sampling entails search for where the seminal paper itself has been cited in other papers

Combinations	<ul style="list-style-type: none"> ○ Combinations of the above approaches can be used to best fit the purpose of the synthesis. ○ For example, criterion sampling could be combined with purposeful random sampling
--------------	---

Example: Sampling for a multi-agency UN evaluation synthesis on SDG5

This was a synthesis of UN evaluations on SDG5 by UN Women, UNICEF, UNFPA, UNDP and WFP. The objectives were to map UN system evaluations on SDG5, synthesize the evaluative evidence to highlight barriers and enabling factors, and to develop lessons to help accelerate progress towards SDG5 by 2030.

Initial searches in 54 UN databases returned 3,150 evaluations. These were screened against the eligibility criteria, which resulted in a universe of 619 relevant evaluations. A target sample size of 300 was selected based on feasibility, timeframe and desire to reflect insights from as many evaluations as possible without posing a risk to the analysis. The universe of 619 evaluations was coded according to pre-determined 'attributes' (such as evaluation agency, type, geographic scope, SDG 5 target, quality assurance rating and year completed).

The review team then applied a sampling strategy that combined stratified purposeful and random sampling. The strategy was designed to maintain representation across UN agencies, regions, and evaluation types.

- Global level evaluations were included in the sample in their entirety, because they focus on multiple contexts, explore a programmatic theme in great depth and may provide more information on broader approaches such as coordination and partnerships.
- The remaining evaluations were proportionately and randomly selected from the sample frame across geographic regions and evaluation types.
- Firstly, a sample target was set for each geographic region that was proportionate to the representation of that region in the universe of evaluations.
- Within each of these regional 'strata', separate targets were set for different evaluation types (e.g., Country Portfolio, Institutional effectiveness, Project/Programme/Thematic evaluations). Again, the sample targets for each evaluation type were proportionate to the representation of these evaluation types in that regional stratum.
- Evaluations were then randomly selected to achieve the stated targets within each stratum.
- To ensure representation across agencies, it was decided that each agency from which an evaluation has been identified would have at least one evaluation report in the sample. To accomplish this, following the random selection of evaluations, the sample was reviewed against the list of agencies in the sample frame. After conducting this step, it was found that all agencies were represented without further action being taken. If required, one evaluation would have been randomly chosen and replaced with another randomly chosen evaluation within the same region and evaluation type.

Useful resources

Cochrane Effective Practice and Organisation of Care (2017). EPOC Qualitative Evidence Syntheses guidance on when to sample and how to develop a purposive sampling frame.

https://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/Resources-for-authors2017/qes_guidance_on_sampling.pdf

Ames, H., Glenton, C. & Lewin, S. (2019). Purposive sampling in a qualitative evidence synthesis: a worked example from a synthesis on parental perceptions of vaccination communication. *BMC Med Res Methodol* 19, 26. <https://doi.org/10.1186/s12874-019-0665-4>

Suri, H. (2011). Purposeful Sampling in Qualitative Research Synthesis, *Qualitative Research Journal*, 11, 2. <https://doi.org/10.3316/QRJ1102063>

Video: Sampling primary studies for inclusion in a qualitative evidence synthesis

<https://www.youtube.com/watch?v=NnCfTNPZUS0>

13. Quality appraisal

13.1. What is the purpose of quality appraisal in a synthesis?

When conducting a synthesis, it is important to ensure that the evaluations and research included are of a certain quality. Having confidence in the findings of individual evaluations and research reports included in a synthesis ensures confidence in the findings of the exercise overall.

13.2. How is the quality of evaluations and research studies appraised?

For UN evaluations, quality can be assessed in terms of: the report structure; the evaluation purpose, objectives, scope, questions and criteria; the evaluation methodology; the coherence and validity of findings and whether they address the evaluation questions and criteria; the relevance of recommendations to the evaluation purpose and whether they are explicit, relevant and actionable; the incorporation of gender and human rights considerations and perspectives in the design of the evaluation, in the analysis, and in the reporting. UNEG has published a quality checklist for evaluations, and many other UN evaluation offices have their own approaches based on similar principles. In many cases this quality appraisal results in an overall rating or benchmarking.

There are many available tools for appraising the quality of other types of evidence, and these differ by research design. They assess different aspects of a study, including the conceptual framework and its relevance; methodological clarity and transparency; validity, reliability and consistency; clarity of reporting/ findings; and the extent to which the authors critically engage with the literature.

13.3. When does quality appraisal occur in the synthesis process?

Generally, a quality appraisal is conducted after the screening stage when the evaluations or studies to be included in the synthesis are known. Report quality may be reassessed during the data extraction or coding phase if new details are observed during the more in-depth review of reports. In some cases, the quality appraisal could occur exclusively as part of the data extraction or coding phase.

13.4. What are the key considerations?

There are several key considerations when planning and conducting the quality appraisal:

- Will all evaluations and studies be included in the synthesis, regardless of their rating? This consideration may be necessary when the universe of evaluations is small, or where there are few higher quality evaluations available to synthesize.
- Will evaluations or studies below a certain quality rating be excluded from the synthesis? Often, only satisfactory or higher-rated evaluations are included in a synthesis. This helps to ensure confidence in the evidence being synthesized and can also be a means of ensuring that the universe of evaluations is manageable.

- Will evaluations need to be quality appraised as part of the synthesis process, or can existing quality ratings be used? Often, UN evaluations are already rated, and therefore further quality appraisal is not needed. In some cases, the universe of evaluations may include non-rated evaluations, and these may need to be appraised using an appropriate checklist. In certain circumstances, it may be acceptable to exclude non-rated evaluations from a synthesis (e.g. when the universe of evaluations is already large, or where there is limited time to conduct the synthesis).

13.5. Quality appraisal for single agency syntheses

It is appropriate to draw on existing quality ratings for evaluations in single agency syntheses. Many evaluations will already be rated, and further appraisal is not necessary, helping to speed up and simplify this step. In instances where the universe of evaluations includes non-rated evaluations, consideration should be given to rating these using the agency-specific rating tool.

13.6. Quality appraisal for multi-agency syntheses

13.6.1 Ensuring the equivalence of UN agency evaluation rating scales

UN agency quality appraisal tools are based on common underlying principles and appraise common aspects of evaluations. However, the rating categories and scales used by different agencies vary. Therefore, it is necessary to develop a framework that aligns the different UN agency ratings. Existing multi-agency UN evaluation syntheses provide examples of how this can be done (see box x).

Box X: Aligning quality appraisal frameworks in the SDG5 Evaluation Synthesis

The review team developed a framework aligning the rating scales of participating agencies. This included UNICEF's Global Evaluation Reports Oversight System (GEROS), UN Women's Global Evaluation Report Assessment and Analysis System (GERAAS), WFP's Evaluation Quality Assurance System (EQAS), UNDP's Evaluation Quality Assessment, and UNFPA's Evaluation Quality Assurance and Assessment (EQAA). In this synthesis, only evaluations rating with satisfactory, mostly satisfactory, or good across these agency scales were eligible for inclusion.

Table x. UN agency evaluation quality assurance system (Table 1 in the report annex).

CATEGORY RATINGS	UNICEF GEROS	UN Women GERAAS	WFP EQAS	UNDP Evaluation Quality Assessment	UNFPA EQAA
	Highly Satisfactory	Very Good	Highly Satisfactory	Highly Satisfactory	Very Good
	Satisfactory	Good	Satisfactory	Satisfactory	Good
				Mostly Satisfactory	
	Fair	Fair	Partially Satisfactory	Mostly Unsatisfactory	Fair
	Unsatisfactory	Unsatisfactory	Unsatisfactory	Unsatisfactory	Unsatisfactory
				Highly Unsatisfactory	

13.6.2 If evaluations are unrated

Unrated evaluations should be appraised using an existing tool. The UNEG Checklist is one option, or it may be possible to adapt and use an external quality rating scale (as illustrated by the SDG 6 Evaluation Synthesis in Box Y).

13.7. Quality appraisal in syntheses of UN evaluations and other types of evidence

Appropriate quality appraisal tools can be used when including other types of evidence in a synthesis. Existing multi-agency syntheses provide examples of how quality appraisal is treated in syntheses including both UN evaluations and other types of evidence (see boxes y and z). Often, the quality appraisal tools are specific to study type and research design. Some suggested quality appraisal tools are listed below.

Box Y: Aligning quality appraisal frameworks in the SDG6 Evaluation Synthesis

This review team also developed a framework aligning the rating scales of participating agencies. This included UNICEF's Global Evaluation Reports Oversight System (GEROS), UNEP's Evaluation Criteria, and UNDP's Evaluation Quality Assessment. This synthesis also included evaluations published by bi- and multi-lateral agencies external to the UN. Therefore, if an evaluation had no equivalent quality rating or no rating system by the implementing agency, a modified checklist developed by BOND was used to rate the quality. Evaluations with a satisfactory or mostly satisfactory rating on UN scales, or a 'pass' rating on the BOND scale were eligible for inclusion.

Table x: Summary of quality assurance measures (Table 5 in the report)

	UNICEF GEROS	UNEP	UNDP	BOND	DECISION
CATEGORY RATINGS	Highly Satisfactory	Highly Satisfactory	Highly Satisfactory	PASS	Eligible for inclusion ✓
	Satisfactory	Satisfactory Mostly satisfactory	Satisfactory Mostly satisfactory		
	Fair	Mostly unsatisfactory	Mostly unsatisfactory	FAIL	Not eligible ✗
	Unsatisfactory	Unsatisfactory Highly unsatisfactory	Unsatisfactory Highly unsatisfactory		

UNICEF (2021). *Evaluation Synthesis of United Nations System and Development Bank Work Towards SDG 6*. New York: UNICEF

Box z: Appraising the quality of UN evaluations and external impact evaluations in a mixed-method synthesis of evidence on the Partnership Pillar of the Sustainable Development Goals

This synthesis included both impact evaluations published in academic repositories and evaluations by UN and other bi- and multi-lateral partners. The impact evaluations were subjected to a ‘risk of bias’ assessment, to determine the rigor of quantitative studies assessing the impact of interventions. All impact evaluations were eligible for inclusion regardless of results of the risk of bias assessment. The quality of evaluations by UN and other bi- and multi-lateral agencies was appraised using a tool that combined and adapted a list of quality indicators from several UN evaluation quality appraisal tools (UNICEF’s GEROS, UNDP’s Evaluation Quality Assessment, UNEP’s evaluation criteria, UNFPA’s EQAA, and WFP’s EQAS). Only those that scored satisfactory or higher were eligible for inclusion.

De Hoop et al., (2023). What works to accelerate progress on the Partnership Pillar of the Sustainable Development Goals: A synthesis of evaluative evidence. New York: The Global SDG Synthesis Coalition

Useful resources

Quality appraisal of UN Evaluations

United Nations Evaluation Group (UNEG). (2010). *UNEG quality checklist for evaluation reports*.
<https://www.unevaluation.org/document/download/853>

Quality appraisal of other types of evidence

The Joanna Briggs Institutes provides several useful resources for the appraisal of external research studies of different research designs: <https://jbi.global/critical-appraisal-tools>

The Critical Appraisal Skills Programme also provides several checklists for the appraisal of different research designs: <https://casp-uk.net/casp-tools-checklists/>

The International Development Research Centre (IDCR) created a Research Quality Plus (RQ+) Assessment Instrument which provides a framework and practical guidelines for assessing the quality of research for development. https://idrc-crddi.ca/sites/default/files/sp/Documents%20EN/idrc_rq_assessment_instrument_september_2017.pdf

Risk of bias tools

Cochrane Collaboration’s risk of bias tool for randomized controlled trials (RoB 2):
<https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials>

Cochrane Collaborations risk of bias tool for non-randomized studies of interventions (ROBINS-I):
<https://methods.cochrane.org/bias/risk-bias-non-randomized-studies-interventions>

14. Coding and extracting data

14.1. What is data extraction?

‘Data’ is information from or attributes of an evaluation or study that are useful for answering synthesis questions. Coding (or data extraction) means collecting this data from evaluations or studies for use in the analysis or synthesis. Data extraction should be based on a pre-defined plan that is outlined in the inception report or protocol. The approaches to data extraction for qualitative and quantitative data will vary.²⁴

14.2. Qualitative data extraction

Qualitative data in UN evaluations and studies derives from interviews, focus group discussions, observations and document analysis. UN evaluations typically take a mixed methods approach, triangulating both qualitative and quantitative data (which can include programme financial data, monitoring data and survey results) to provide insights into the relevance, effectiveness, efficiency or sustainability of a programme, project, or intervention. The resulting findings are presented as text and can be subject to qualitative data extraction.

Qualitative data extraction is commonly used for syntheses that set out to answer questions relating to how or why a programme, project or intervention works. In many cases the extraction and analysis focus on the facilitators and barriers to success or progress, which can refer to implementation features or contextual issues. However, given the mixed-methods nature of evaluations, qualitative data extraction can also be used for syntheses that set out to answer questions on what programmes, projects or interventions work. Based upon a mixed-method approach, evaluations may note success in terms of outcomes, outputs and results, improvements in the implementation process, or interventions that led to clear policy or legal changes, for example.

Like primary qualitative research, qualitative data extraction entails a flexible and iterative approach.²⁵ Part of this iterative approach may involve both deductive and inductive approaches to extraction. This may begin with a ‘top down’ approach based upon a pre-determined framework of concepts or categories (e.g. a conceptual framework), which will be applied to evaluations and studies. There may then be a process of reading, comparing and re-reading evaluations and studies, through which deeper, richer or extended concepts or categories may emerge as part of a ‘bottom-up’ inductive process.

14.3. Quantitative data extraction

Quantitative data in UN evaluations involves the use of numerical information and statistical methods to assess the performance, results, outcomes and impact of programmes, projects or interventions.

²⁴ Bakrania, S. (2020). Methodological Briefs on Evidence Synthesis: Brief 4 – Collating and analysing studies for synthesis, Innocenti Research Brief 2020-04. Florence: United Nations Children’s Fund Office of Research – Innocenti. <https://www.unicef-irc.org/publications/pdf/IRB%202020-04.pdf>

²⁵ Noyes J & Lewin S. (2011). Chapter 5: Extracting qualitative evidence. In: Noyes J, Booth A, Hannes K, Harden A, Harris J, Lewin S, Lockwood C (editors), *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions*. Version 1 (updated August 2011). Cochrane Collaboration Qualitative Methods Group. <http://cqrmg.cochrane.org/supplemental-handbook-guidance>

Specific types of data may include outcome indicators, output indicators, baseline and endline surveys, and impact assessments to assess causal impacts of interventions using a counterfactual.

Compared to the iterative nature of qualitative data extraction, quantitative data extraction is more 'fixed' and linear, focusing on descriptive data on what was done in the study, as well as numerical data that reports the measure of effect of the programme, project or intervention. Key items for data extraction should be defined in advance in a data collection form. This aggregative data can then be used for quantitative synthesis, involving meta-analysis or narrative synthesis, for example.

14.4. Data extraction forms

For all types of data extraction, it is necessary to develop a data extraction framework. Flexibility should be allowed for some types of iterative qualitative data extraction, where additional categories may be added as part of the process. The framework should be fixed for quantitative data extraction.

The type of data to be extracted from evaluations or studies will usually include:

- Evaluation or study title
- Bibliographic information for research studies
- Evaluation type or study research design
- Implementing agency
- Geographic scope/setting/context
- Theme/Intervention/Programme
- SDG targets and/or related SDG
- Quality appraisal rating
- Key findings that respond to the synthesis questions (for qualitative synthesis)
- Measures and size of effects (for quantitative synthesis)

Primary data gathering, such as interviews with key stakeholders, may be considered to verify and deepen data from component evaluations. These should take place once data extraction has been completed and the synthesis team has had sufficient time to conduct a preliminary analysis. It is important to note that using interviews for synthesis differs from their use in evaluation. During synthesis, interviews are valuable when initial findings are available, as they aid in contextualization rather than gap filling. Tools for data extraction

At a very basic level, data extraction can be conducted in a simple Excel sheet, with rows for each evaluation or study, and columns for entering data. For qualitative synthesis, coding may be conducted in qualitative data analysis software such as NVivo. Dedicated systematic review management software for example, EPPI-Reviewer, has functions for both qualitative and quantitative data extraction and analysis.

Machine learning and artificial intelligence tools for data extraction are a work in progress, and data extraction is still mostly conducted manually.²⁶ There are some publicly available tools that can aid or

²⁶ Li T., Higgins J.P.T. and Deeks J.J. (2023). Chapter 5: Collecting data. In: Higgins JPT et al. (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). www.training.cochrane.org/handbook.

complement but not entirely supplement the data extraction effort. UNDP's Artificial Intelligence for Development Analytics (AIDA) tool can help identify key conclusions or findings from UNDP evaluations, has topic modelling functions that help quickly identify key attributes of evaluations, and an analytical insights function that provides a summary of findings from bodies of evaluations.²⁷

Useful resources

Guidance on qualitative data extraction

Noyes, J. et al. (Eds) Chapter 21: Qualitative evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (2023). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). Cochrane, 2023. www.training.cochrane.org/handbook.

Campbell Collaboration Qualitative Evidence Synthesis Workshop Video:

https://www.youtube.com/watch?v=DA_Zln4U8dU

Guidance on quantitative data extraction

Li T., Higgins J.P.T. and Deeks J.J. (2023). Chapter 5: Collecting data. In: Higgins JPT et al. (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). www.training.cochrane.org/handbook.

²⁷ At the time of writing, AIDA only incorporates UNDP evaluations.

15. Analyzing and synthesizing

14.1 What is the purpose of analyzing and synthesizing?

Analyzing and synthesizing involves organizing, interpreting, and summarizing data or evidence from evaluations or studies to address synthesis questions. The goal is to integrate individual pieces of evidence into a cohesive whole. The methods used depend on the synthesis questions, objectives, and types of evidence included.

14.2 Structuring the analysis

There are several approaches to structuring the analysis or synthesis. One natural option is to organize the evidence around the synthesis questions (see Section 6: *Defining Evaluation Synthesis Questions*), as these questions underpin decisions about the types of evidence included. The analysis can be divided into components that address each synthesis question, which may involve analyzing or triangulating different subsets of extracted evidence to answer specific questions.

Alternatively, the analysis can follow a conceptual framework or theory of change, as outlined in the protocol or inception report. This approach, often referred to as *framework synthesis*, uses elements of the framework or theory of change as categories or themes to organize the synthesis (see 14.5 on *Qualitative synthesis methods* below). It is particularly useful when testing the validity of a framework or theory of change, helping to assess whether its logic is supported by the available evidence.

14.3 Overview of synthesis methods

There are different methods for answering different types of synthesis questions. Section 6 of this guidance on ‘Defining evaluation synthesis questions’ introduces the distinction between questions on ‘what works’ and ‘how or why something works’. Answering these two different types of questions entails the use of different synthesis methods. In most cases:

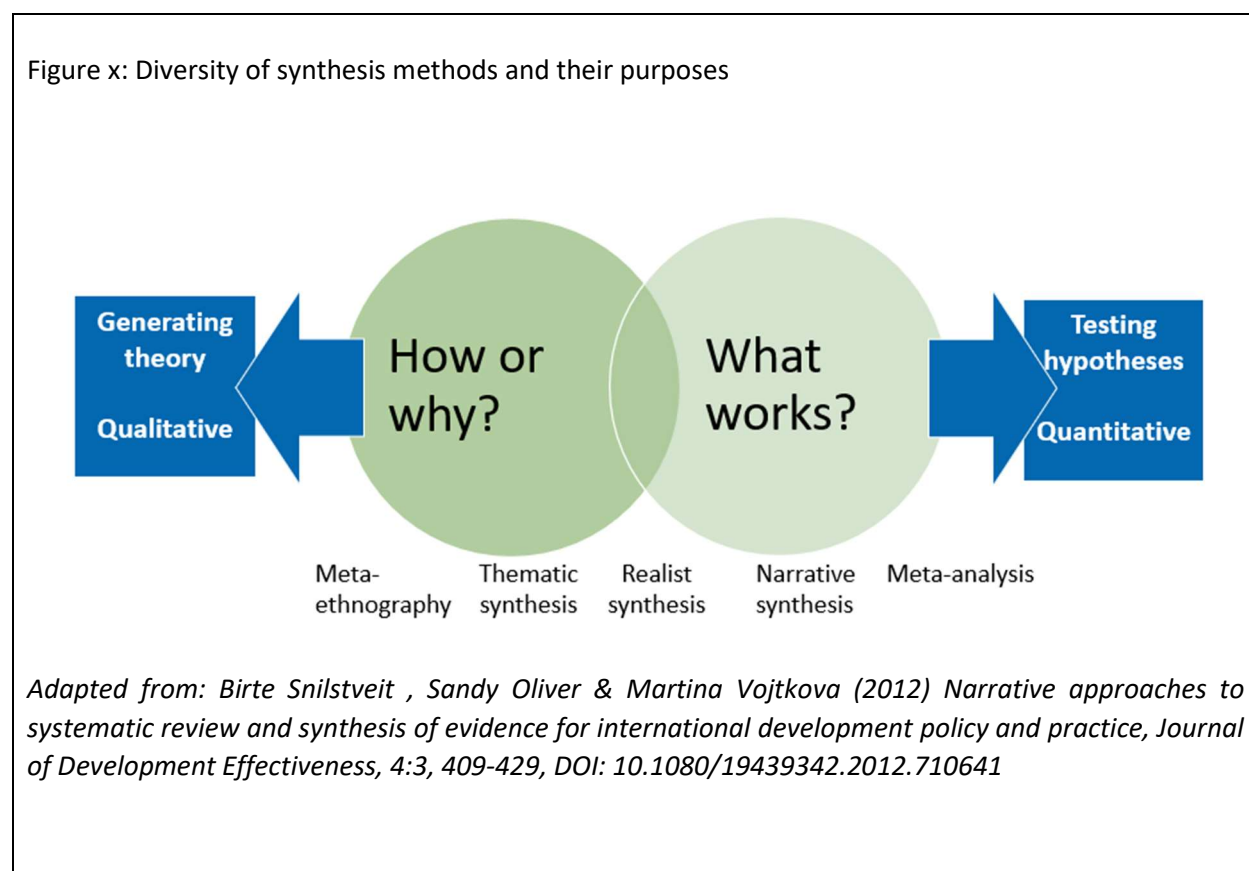
Answering ‘what works’ type questions involves testing hypotheses about the effects of an intervention to help determine what kinds of intervention works. Methods used to answer these questions are sometimes referred to as ‘aggregative’, where (usually quantitative) data from experimental or quasi-experimental impact evaluations is combined or pooled to determine the average or aggregate effect of interventions. This data can then be reported statistically through ‘meta-analysis’ or narratively through ‘narrative synthesis’.²⁸ To some extent, qualitative evidence from UN evaluations may also be used to answer questions on effectiveness, particularly specific findings in evaluations that speak to the OECD-DAC evaluation criteria.

Answering ‘how or why something works’ type questions involves adding contextual and descriptive analysis to generate theories or insights about why certain interventions are effective or not. Methods used to answer these questions are sometimes referred to as ‘configurative’, whereby (usually qualitative) data on design, implementation or contextual factors is collated from multiple studies or evaluations and

²⁸ Bakrania, S. (2020). Methodological Briefs on Evidence Synthesis: Brief 4 – Collating and analysing studies for synthesis, Innocenti Research Brief 2020-04. Florence: United Nations Children’s Fund Office of Research – Innocenti. <https://www.unicef-irc.org/publications/pdf/IRB%202020-04.pdf>

configured or interpreted using qualitative data analysis techniques to infer a theory or to build an argument.^{29,30}

The diversity of methods, and their different purposes are illustrated in figure x below.



14.4 Synthesis methods to answer ‘what works’ type questions

14.4.1 Meta analysis

Meta-analysis is a common statistical method used in systematic reviews combining numerical evidence from multiple experimental and quasi-experimental impact evaluations to produce an overall summary of knowledge on a given topic. The term ‘meta-analysis’ has a very specific meaning in synthesis, and should not be confused or conflated with the terms ‘meta-evaluation’ or ‘meta-synthesis’, which are sometimes used within the UN evaluation system to denote ‘an evaluation of evaluations’ (see Section 2.1). This approach is less common in UN evaluation syntheses because quantitative impact evaluations are not commonly produced by most UN evaluation agencies.

²⁹ ibid

³⁰ Birte Snilstveit , Sandy Oliver & Martina Vojtkova (2012) Narrative approaches to systematic review and synthesis of evidence for international development policy and practice, *Journal of Development Effectiveness*, 4:3, 409-429, DOI: 10.1080/19439342.2012.710641

In this process, the findings of impact evaluations are transformed into metrics known as 'effect sizes,' which are then aggregated to calculate an overall average effect size. The effect size quantifies the magnitude of an intervention's impact by comparing two groups—typically, one receiving the intervention and the other receiving no intervention or an alternative intervention.^{31, 32}

Meta-analysis combines impact evaluations with identical interventions and outcomes. This allows effects sizes to be aggregated across countries and contexts. Doing so increases sample size, improves precision of impact estimates, and allows the influence of moderating factors to be examined.

When conducting meta-analysis, evaluators use several steps take to create the necessary data frame and ensure it is robust.

- **Coding and extracting data:** Two separate reviewers typically use a data extraction form with discrepancies reconciled by a third party.³³ Extractions include the effect size reported in the study, details on calculations (for example, whether a whole sample or sub-sample effect size is given), sample sizes (for treatment and comparison groups), standard deviations of outcome variables, test statistics (such as t-test, F-test, p-values), and standard errors of effect estimates.
- **Further coding:** can include analysis method, type of comparison group, wider information on the design of the intervention, outcomes and study. For example, intervention/outcome sub-categories, and potential moderators (such as study design, scale, urban/rural, demographic characteristics, implementing agency type, duration, intensity).
- **Standardization of effect sizes:** is necessary to allow comparisons across studies with different scales or measures. Effect sizes are converted into a common metric to ensure consistency and alignment, such as standardized mean difference for continuous outcomes, odds ratios for binary outcomes, and correlation coefficients for relationships between variables. Effect sizes are also adjusted to ensure that they are coded in the same direction (e.g. that higher values show positive effects or improvements)
- **Critical appraisal and risk of bias:** Evaluators often apply a critical appraisal or risk of bias tool to assess the trustworthiness, rigor and credibility of primary studies included in meta-analysis. These tools examine the extent to which a study's results may be influenced by systemic errors or biases, including selection bias, performance bias and reporting bias. This leads to an overall risk of bias score which is used to determine confidence in findings. In many case, low confidence studies will be excluded from the synthesis.
- **Assessing the strength of bodies of evidence:** The final step in meta-analysis is to assess the overall strength of the evidence base included. Tools to assess the strength of evidence are used to help evaluators, researchers, policymakers and practitioners make informed decisions by

³¹ Littell, Julia H., Jacqueline Corcoran and Vijayan Pillai (2008). *Systematic Reviews and Meta-analysis*, Oxford University Press, Oxford, 2008. Available at:

<http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195326543.001.0001/acprof-9780195326543>

³² Coe, Robert. (2002), 'It's the Effect Size, Stupid: What effect size is and why it is important', Paper presented at the British Educational Research Association annual conference, Exeter, 12–14 September 2002. Available at: <http://www.cem.org/attachments/ebe/ESguide.pdf>

³³ Reviewers and the third party usually pilot the data extraction tool drawing a random sample of impact evaluations with the aims of achieving 90% alignment before finalizing the tool.

assessing the confidence they can place in the aggregated findings across a body of evidence. This is distinct from the critical appraisal step.

14.1.2 Narrative synthesis

Narrative synthesis is a method used to aggregate and explain the findings of multiple impact studies through text, rather than through statistical approaches. This method is often used when meta-analysis is not feasible – e.g. when there are too few identical studies, or when included studies are considered too diverse in terms of interventions or outcomes. This approach is less common in UN evaluation syntheses because quantitative impact evaluations or outcome level evaluations are not commonly produced by most UN evaluation agencies.

Narrative synthesis can include the following steps:³⁴

- **Develop a theory:** on how the intervention works, for whom and under what circumstances. This may be in the form of a theory of change or conceptual framework, which can be detailed in the inception report protocol (see section 8).
- **Undertake a preliminary synthesis:** organizing, summarizing or describing the data or evidence extracted from included evaluations or studies to identify patterns. This may involve collating descriptive information about each study or evaluation (e.g. on type of intervention, geographical setting, type of evaluation), grouping included studies according to theme, sector or context, tabulating effect size estimates and ensuring some form of alignment or consistency in how this data is tabulated (e.g. a common measure on the size and direction of effect) and undertaking a qualitative or thematic analysis to report the findings (see below).
- **Explore relationships between and within studies:** Patterns or factors may emerge during the preliminary synthesis that explain differences in outcomes, such as target populations, contexts or differences in study design.
- **Assess the strength of bodies of evidence:** The final step in meta-analysis is to assess the overall strength of the evidence base included. Tools to assess the strength of evidence are used to help evaluators, researchers, policymakers and practitioners make informed decisions by assessing the confidence they can place in the aggregated findings across a body of evidence. This is distinct from the critical appraisal step. In narrative synthesis, this may involve apply rules or criteria to define weak, moderate or good evidence for different outcomes being studies.

14.5 Qualitative synthesis methods to answer ‘how and why’ type questions

There are a variety of qualitative synthesis methods used for answering ‘how or why something works’ type questions. Qualitative synthesis is the process of combining and synthesizing evidence from multiple qualitative evaluations and studies to derive overarching insights, themes or concepts. Qualitative synthesis is particularly useful for answering ‘how or why’ type questions because it can enhance the understanding of complex phenomena, including experiences, perspectives, behaviors in the context of development interventions, as well as an understanding of contextual, design and implementation issues

³⁴ Centre for Reviews and Dissemination. (2008). CRD’s guidance for undertaking reviews in health care. York: University of York. Available at: [Systematic Reviews: CRD's guidance for undertaking reviews in health care](#)

that influence the effectiveness of interventions.³⁵ This approach is commonly used in UN evaluation syntheses because of the qualitative nature of most UN agency evaluations.

Common approaches are labelled ‘thematic synthesis’, ‘framework synthesis’ or ‘meta-ethnography’. While the stated purpose and outcome of these different methods vary, they all draw from common qualitative data analysis or thematic analysis approaches that entail a process of constant comparison between evaluations or studies. As part of this process, evaluations or studies are coded or categorized into analytical themes. The coding process can be deductive (a top-down approach based on an existing theory or framework) or inductive (a bottom-up approach where themes emerge through comparison and analysis), or a combination of both. These analytical themes are then arranged into a theory or a narrative to answer the synthesis question/s.

Framework synthesis uses existing theoretical or conceptual frameworks to guide the synthesis of qualitative data. As part of the process, data from evaluations or qualitative studies is mapped onto a pre-defined framework, which provides a deductive structure or ‘scaffold’ for analysis and comparison. It can be an effective approach when the synthesis question or the objective of the synthesis relates to a policy framework.³⁶

Meta-ethnography entails examining key concepts and theories within and between studies, comparing these to highlight similarities and differences, and then organizing them into conceptual or theoretical categories to develop a theoretical or conceptual framework.³⁷ This approach is highly inductive, interpretive and abstractive, seeking to develop a deep understanding of a phenomenon or to generate new theories through the synthesis process. It is useful for exploring complex phenomena and to refine theoretical understanding.

14.6 Mixed method and realist synthesis to explore ‘what works for whom and in what circumstances’

Mixed-methods synthesis systematically combines qualitative and quantitative evidence to provide a comprehensive understanding of ‘what works’ and ‘how or why something works’. In this way, mixed methods syntheses can provide useful insights into complex interventions to help policy makers and programme designers. This includes providing insights on other factors that influence intervention effectiveness, target populations who may benefit more or less from interventions, and the influence of contextual or implementation issues. This involves integrating findings from diverse types of evidence to address complex synthesis questions and to inform decision-making. Mixed methods syntheses which integrate UN agency evaluations and other types of evidence – e.g. externally published research studies

³⁵ Noyes J, Booth A, Cargo M, Flemming K, Harden A, Harris J, Garside R, Hannes K, Pantoja T, Thomas J. Chapter 21: Qualitative evidence [last updated October 2019]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5. Cochrane, 2024. Available from www.training.cochrane.org/handbook.

³⁶ Brunton G, Booth A, Carroll C. Chapter 9. Framework Synthesis. Draft version (March 2024) for inclusion in: Noyes J, Harden A, editor(s). *Cochrane-Campbell Handbook for Qualitative Evidence Synthesis*, Version 1. London: Cochrane

³⁷ Cahill, Mairead, et al., ‘Qualitative Synthesis: A guide to conducting a meta-ethnography’, *British Journal of Occupational Therapy*, vol. 81, no. 3, 2018, pp. 129–137. Available at: <https://doi.org/10.1177/0308022617745016>

- are increasingly common in UN evaluation synthesis activities, as evidenced by the work of the Global SDG Synthesis Coalition.

In this approach, the integration of evidence from quantitative and qualitative synthesis is fundamental (e.g. combining a narrative synthesis of quantitative impact evaluations with a thematic synthesis of qualitative UN evaluations). There are several approaches:³⁸

- **Assimilation:** Quantitative and qualitative evidence on a similar topic can address the same synthesis question/s, and so they are synthesized together. This may involve transforming findings so that they can be merged (e.g. transforming qualitative data into quantitative data or vice versa)
- **Comparison:** Comparing the results from independent quantitative and qualitative syntheses. This recognizes the distinct methods for qualitative and quantitative synthesis, whilst allowing for one type of synthesis to help explain the findings from another. For example, a meta-analysis or narrative synthesis of impact evaluations may provide findings on the impact of effectiveness of interventions, while a qualitative synthesis of UN evaluations will provide further details on how context, or the design or implementation of interventions influences effectiveness.
- **Connection:** The results of one type of synthesis are used to inform the focus and conduct of another. The quantitative and qualitative syntheses are conducted separately, but there is an ongoing discussion of emerging findings between the bodies of evidence to inform further questioning or analysis. Alternatively, a qualitative and quantitative synthesis may proceed in a sequential manner, with the results of the first synthesis informing the focus and scope of the latter. For example, a quantitative synthesis may suggest that outcomes or results vary for different populations. A subsequent qualitative synthesis of UN evaluations could explore the variation between groups. Alternatively, a qualitative synthesis may suggest alternative priorities, outcomes or results that matter to different population groups, a subsequent qualitative synthesis then explore the effectiveness of interventions on these prioritized outcomes.

Realist synthesis is a specific mixed method approach and can also be seen as a method to answer both 'what works' and 'how and why something works' type synthesis questions. It is a method for studying complex interventions, focusing on the contextual factors and mechanisms that mediate the relationship between interventions and outcomes, making it particularly useful in the early stages of implementation. By examining barriers, enablers, and theories of change, this approach iteratively tests context-mechanism-outcome (CMO) configurations to refine programme design. A key advantage of realist synthesis is its explanatory nature, helping identify when, why, and how interventions are likely to succeed across different contexts and for different groups. Another advantage is that it is inclusive and flexible in its approach to including different types of evidence, ranging from quantitative impact evaluations to interpretive and qualitative evidence to policy documents that highlight the intended design and delivery of interventions.³⁹

³⁸

³⁹ Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). *Realist review—a new method of systematic review designed for complex policy interventions*. *Journal of Health Services Research & Policy*, **10**(1), 21–34.

An explicit realist synthesis approach is rare in UN evaluation synthesis. It requires time, resources and expertise. However, elements of the realist approach, such as developing synthesis questions to explore contextual, design and implementation factors that influence effectiveness, are increasingly being adopted.⁴⁰

The six stages of realist synthesis begin with defining synthesis questions to prioritize programme theories and hypotheses, based around CMO configurations. The search strategy follows, identifying relevant sources including a range of evidence types and applying tools like SPIDER to establish the eligibility criteria (see section 9). Searching and screening may develop in an iterative manner (see sections 10 and 11). Next, evidence appraisal entails assessing whether evidence helps understand the context, mechanisms or outcomes relevant to the synthesis question. Data extraction focuses on capturing patterns of how specific contexts trigger mechanisms to produce outcomes. Model refinement includes comparative and iterative refinements of programme theories (defined in CMO configurations) to strengthen an overarching theory of change. Finally, the synthesis must be concise, engaging, and policy-oriented, emphasizing the importance of context, evidence gaps, and lessons for effective intervention design.⁴¹

Useful resources

Guidance on meta-analysis

Littell, Julia H., Jacqueline Corcoran and Vijayan Pillai (2008). *Systematic Reviews and Meta-analysis*, Oxford University Press, Oxford, 2008. Available at:

<http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195326543.001.0001/acprof-9780195326543>

Deeks JJ, Higgins JPT, Altman DG, McKenzie JE, Veroniki AA (editors). Chapter 10: Chapter 10: Analysing data and undertaking meta-analyses [last updated November 2024]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5. Cochrane, 2024. Available from www.training.cochrane.org/handbook.

Guidance on narrative synthesis

Centre for Reviews and Dissemination. (2008). *CRD's guidance for undertaking reviews in health care*. York: University of York. Available at: [Systematic Reviews: CRD's guidance for undertaking reviews in health care](http://www.crd.york.ac.uk/crd/CRDguidance.htm)

McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods [last updated October 2019]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5. Cochrane, 2024. Available from www.training.cochrane.org/handbook.

Cochrane Training video on 'Definition and use of 'narrative synthesis': <https://www.youtube.com/watch?v=GXkJA4RGyVw>

Overall guidance on qualitative evidence synthesis

⁴⁰ Ibid.

⁴¹ Ibid.

Cochrane-Campbell Handbook on Qualitative Evidence Synthesis (2023)
https://training.cochrane.org/cochrane-campbell-handbook-qualitative-evidence-synthesis?utm_source=chatgpt.com

Guide to qualitative evidence synthesis: evidence-informed policy-making using research in the EVIPNET framework. Copenhagen: WHO Regional Office for Europe; 2021
<https://iris.who.int/bitstream/handle/10665/340807/WHO-EURO-2021-2272-42027-57819-eng.pdf?sequence=1>

Guidance on framework synthesis

Brunton G, Booth A, Carroll C. Chapter 9. Framework Synthesis. Draft version (March 2024) for inclusion in: Noyes J, Harden A, editor(s). Cochrane-Campbell Handbook for Qualitative Evidence Synthesis, Version 1. London: Cochrane <https://training.cochrane.org/cochrane-campbell-handbook-qualitative-evidence-synthesis/qeschapter9frav0250324>

Guidance on meta-ethnography

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography*. SAGE Publications, Inc.,
<https://doi.org/10.4135/9781412985000>

Garside R, France E, Noyes J. Chapter 11: Conducting a meta-ethnography. Draft version (September 2023) for inclusion in: Noyes J, Harden A, editor(s). Cochrane-Campbell Handbook for Qualitative Evidence Synthesis, Version 1. O. London: Cochrane <https://training.cochrane.org/cochrane-campbell-handbook-qualitative-evidence-synthesis/qeschapter11metv0231023>

Guidance on realist synthesis

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). *Realist review—a new method of systematic review designed for complex policy interventions*. Journal of Health Services Research & Policy, **10**(1), 21–34.

Better evaluation realist synthesis: <https://www.betterevaluation.org/methods-approaches/methods/realist-synthesis>

Introduction to realist review (video): https://www.youtube.com/watch?v=t_Z1gycJHzc

Guidance on mixed method synthesis

Hong, Q. N., Rees, R., Sutcliffe, K., & Thomas, J. (2020). Variations of mixed methods reviews approaches: A case study. *Research Synthesis Methods*, **11**(6), 795–811.
<https://doi.org/10.1002/jrsm.1437>

Stern, C., Lizarondo, L., Carrier, J., Godfrey, C., Rieger, K., Salmond, S., Apóstolo, J., Kirkpatrick, P., Loveday, H. (2020). Methodological guidance for the conduct of mixed methods systematic reviews. *JB I Evidence Synthesis* **18**(10):p 2108-2118, October 2020. DOI: 10.11124/JBISRIR-D-19-00169

16. Reporting the results of a synthesis

16.1. What is the purpose of reporting in a synthesis?

Reporting is the final phase that aims to share the findings, lessons and/or recommendations from a synthesis in an accessible and appropriate manner for target audiences (who will have been identified earlier in the process). While recognising that there will be variation across UN agencies with respect to reporting, this section aims to identify some common features and best practice.

16.2. What are the key considerations for the reporting phase?

This section describes written deliverables such as synthesis reports, although other forms of final deliverables are possible. The reporting phase requires explicit descriptions of all phases in final deliverables for target audiences. Some ideas include:

- **Who is the audience for the final deliverable?** The target audiences should have been identified in inception phase of the synthesis. Decisions on the level of detail, length and the structure of the report should be made accordingly. Target audiences could include member countries, board members, and staff in other UN agencies, who may not want lengthy and detailed reports; whereas researchers from academic institutions or practitioners from civil organizations may wish to see more detail, especially in terms of the conceptual and methodological approach.
- **To what extent does the synthesis methodology need to be described?** Methodological transparency is an underlying principle of evidence synthesis. However, in certain circumstances, and for certain audiences, a detailed methodology section may not be best placed in the synthesis report. In some cases, it may be desirable to briefly describe the methods in the main body of the report and place further details in an annex, or in a supplementary document such as a protocol, which is then published alongside the synthesis report.
- **To what extent do final deliverables need to include recommendations?** If the synthesis is being conducted for accountability purposes, then the report will likely contain recommendations. If it is being undertaken for learning purposes, then the report will likely emphasize lessons.
- **How should final synthesis reports be structured?** The report (and analysis contained within) should clearly answer the synthesis questions and address the synthesis purpose. As such, the findings and lessons responding to each synthesis question may provide the structure for the analytical segment of the report. Further suggestions for structuring synthesis reports are provided below.
- **How can the findings, lessons and recommendations be made accessible?** Consideration should be given to writing clear and explicit Executive Summaries as part of the report, or separate plain language summaries to accompany the main report. Summaries and briefs present the findings, lessons and recommendations in simple and straightforward language without technical vocabulary, jargon or abbreviations. An executive summary or brief could be a stand-alone document that explains complex ideas through easy and understandable structures and expressions. The goal is to help target audiences understand the synthesis findings, lessons and recommendations.¹

- **What additional resources are needed to promote uptake and use of the synthesis?** To share synthesis findings, it may be necessary to add additional resources to support further dissemination of written products, or to conduct a learning session to promote uptake and use of the synthesis. Consideration should be given to who the audience should be, and how they can most effectively be engaged to share synthesis findings and learning.

16.3. How could a synthesis report be structured?

Included here are some general suggestions to consider when synthesis project teams organize their final deliverables.

- **Abstract:** An abstract includes essential information to describe a synthesis exercise in approximately 400 words. An abstract's purpose is to tell a story through data in a succinct way by describing what was performed, without citations, footnotes or quotations. It contains concise descriptions without sacrificing important content. An abstract includes information on the synthesis objectives, questions, eligibility criteria, intervention/topics/thematic areas/SDGs covered, the types of evaluations and research studies considered for inclusion, and the overall findings and lessons.
- **Executive summary:** An executive summary could be read as an independent document and delivered as a separate report from the main deliverable. It could also provide the basis for a separate plain language brief. It provides an overview of a synthesis exercise for target stakeholders who are short on time to digest key information. Best practices guidance states that the executive summary document is recommended to be no more than four pages.⁴
- **Background and context:** This section provides the rationale for the synthesis, explaining the problem (e.g. stalled progress toward a particular strategic outcome or SDG). It explains how the synthesis came to be – e.g. it is being done for accountability or learning purposes, and does it respond to a clearly identified strategic objective? It provides some contextual information for the synthesis, for example, providing an overview of current progress towards an SDG or strategic outcome, or defining the geographic context that the synthesis focuses on. It also clearly identifies the audience or primary users of the synthesis.
- **Objectives and synthesis questions:** Clearly and explicitly define the synthesis rationale or purpose and objectives and questions.
- **Scope:** Clearly define the conceptual framework and eligibility criteria guiding the synthesis. The conceptual frameworks could provide conceptual and/or analytical structures and should clearly define the thematic scope of the synthesis. The eligibility criteria should establish standards to make decisions on inclusion and exclusion of evidence.
- **Methods:** Clearly describe the search strategy, the approach to screening, the sampling methodology (if appropriate), the approach to quality appraisal, the approach to data extraction/coding, and the synthesis methodology. This section should also detail any quality assurance methods used to ensure the consistency of screening and coding between different members of the synthesis team (e.g. to what extent were evaluations and studies double screened or coded). Some readers might be familiar with synthesis methods and interested in a detailed account, while others might be unfamiliar and interested in the substantive findings and lessons. As stated earlier, there may be variation in how detailed the methods section will be in

the final report. Depending upon the audience, and for accessibility purposes, the detailed methodology could be included in an annex, supplementary report, or in the protocol.

- **Evidence base/mapping:** This section should describe the results of searching and screening. It is useful to depict this schematically using a PRISMA diagram.⁴² Subsequently, provide a descriptive analysis on the nature and quantity of evidence included in the synthesis. This can include visual depictions and a simple narrative describing quantities by evaluation types, geographic distribution, SDG/outcome/topic/theme covered, and other cross-sectional or equity theme covered (such as gender, human rights, disabilities and vulnerable communities).
- **Results:** This section includes the results of the synthesis (i.e. the analysis or interpretation of the evidence collated for the synthesis) and is likely to form the bulk of the report. There are different ways of structuring this section, but the synthesis questions should be clearly answered through a combination of findings that coalesce into broader lessons. It may be useful to structure the results section by synthesis question, and to include the lessons that respond to each question, underpinned by the findings from the evidence collated.
- **Overall lessons:** It may be useful to collate common themes or key lessons into one section.
- **Implications for evaluation and research:** It may be useful to provide some reflections on possible avenues for future evaluation and research, based upon the mapping conducted for the synthesis and an assessment of the gaps. It may also be useful to suggest future avenues for synthesis, e.g. deep diving into particular themes or issues that have arisen from the synthesis, and where the mapping for the synthesis shows that there is a demonstrable body of evidence available for synthesis.
- **Implications or recommendations:** If the synthesis is for learning purposes, then it may not be necessary to provide implications for practice or recommendations requiring a formal management response. If the synthesis is for accountability purposes, then implications and recommendations will be necessary to support the interpretation and use of synthesis findings and lessons. It may be useful to clearly indicate the recommendations and implication relevant for different audiences or stakeholders (e.g. policy makers, practitioners, UN agencies, donor entities etc.) In the case of multi-agency syntheses, it is particularly important to discuss and agree on the management response mechanism.
- **References:** The references section should clearly delineate the evaluations and studies included in the synthesis, from other documents used for the background, contextual or conceptual sections of the report.
- **Annexes:** Additional methodological or technical information could be provided in Annexes, such as a detailed methodology, detailed of evaluation or study characteristics (e.g. a bibliographic table with key characteristics or data extracted from each evaluation or study), screening protocols, coding frameworks, Terms of Reference etc.

Box x: Findings, lessons and recommendations. What's the difference?

Finding: A synthesis finding uses evidence from more than one evaluation or study to allow for a factual statement. Findings are developed by analyzing data from evidence, and should address the

⁴² See <http://www.prisma-statement.org/PRISMAStatement/FlowDiagram> for further guidance on how to construct a PRISMA diagram.

synthesis questions and objectives. Following the principles of synthesis, a finding should never be based on evidence from one evaluation or study alone.

Lesson: Generalizations based on evaluation experiences with projects, programs, or policies that abstract from the specific circumstances to broader situations, capturing a shift in our understanding. Frequently, lessons highlight positive or negative factors that affect performance, outcome, and impact to avoid common mistakes, support learning processes and guide action for decision-makers. Synthesis lessons can be developed from one or more finding and address the following:

- Concisely capture the context from which it is derived to clearly address transferability to other contexts
- Include learnings from both successes and failures
- Focus on learning beyond the obvious, or what is commonly known among development practitioners
- Provide evidence to confirm/disconfirm perceived established “good/bad practices”
- Provide evidence on emerging “promising practices” that are innovative and may spark interest, clearly addressing their applicability.

Recommendation: Proposals aimed at enhancing the relevance, coherence, effectiveness, efficiency, impact or sustainability of the subject of the synthesis. Synthesis recommendations should clearly and logically be based on findings and lessons.

Adapted from OECD-DAC. (2022). DAC Network on Development Evaluation: Glossary of Key Terms in Evaluation and Results-Based Management. Paris: OECD DAC. Additional inputs by the UNDP IEO Synthesis and Lessons Team.

Useful resources

Abstracts, executive summaries, briefs and plain language summaries

Cochrane Collaboration. (2011). Writing an abstract. Cochrane handbook for systematic review of Interventions. https://handbook-5-1.cochrane.org/chapter_11/11_8_writing_an_abstract.htm

Better Evaluation/Global Evaluation Initiative (GEI). (n/a). Executive summaries. <https://www.betterevaluation.org/methods-approaches/methods/executive-summaries>

Campbell Collaboration. (n/a). How to write a plain language summary for a Campbell systematic review. https://www.campbellcollaboration.org/images/How_to_write_a_Campbell_PLS.pdf

Pitcher, N., Michell, D. & Hughes C. (2022). Template and guidance for writing a Cochrane plain language summary. In Higgins, J. & Thomas J. (Eds.). Cochrane Handbook for Systematic Reviews of Interventions. <https://training.cochrane.org/handbook/current/guidance-writing-cochrane-plain-language-summary.pdf>

Pitcher, N., Michell, D. & Hughes C. (2022). Supplementary material: Guidance for writing a Cochrane Plain language summary. In Higgins, J. & Thomas J. (Eds.). Cochrane Handbook for Systematic Reviews of Interventions. <https://training.cochrane.org/handbook/current/chapter-iii-s2-supplementary-material>

International Labour Organization (ILO). Checklist 8: Preparing the evaluation summary for projects. https://www.ilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_166361.pdf.

Joanna Briggs Institute (JBI). (2022). JBI manual for evidence synthesis. <https://jbi-global-wiki.refined.site/space/MANUAL/4687213/10.3.3+Abstract>

On reporting the results of systematic searches and screening

Page, M.J. et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, 71. <http://www.prisma-statement.org/PRISMAStatement/FlowDiagram>

On findings, lessons and recommendations

OECD-DAC. (2022). DAC Network on Development Evaluation: Glossary of Key Terms in Evaluation and Results-Based Management. Paris: OECD DAC. <https://www.oecd.org/dac/evaluation/glossaryofkeytermsinevaluationandresultsbasedmanagement.htm>

MacFarlan, A. (n/a). Lessons learnt. Better Evaluation/Global Evaluation Initiative (GEI). <https://www.betterevaluation.org/methods-approaches/methods/lessons-learnt#:~:text=Lessons%20learnt%20can%20take%20the,experiences%20in%20undertaking%20the%20evaluation>.

Mountain, A. (2014). Recommendations in evaluation. Better Evaluation/Global Evaluation Initiative (GEI). <https://www.betterevaluation.org/tools-resources/recommendations-evaluation>.

Annex 1 List of UN Agency evaluation databases

Table x: UN agency databases

UN agency/UNEG members	URL
ECLAC (United Nations Economic Commission for Latin America and the Caribbean)	https://www.cepal.org/en/publications
ESCAP (United Nations Economic and Social Commission for Asia Pacific)	https://www.unescap.org/kp?f%5B0%5D=kp_sdg_areas%3A6
ESCWA (United Nations Economic and Social Commission for Western Asia)	https://www.unescwa.org/resources?f%5B0%5D=publication_type%3A165 & https://archive.unescwa.org/unbis/evaluation
FAO	https://www.fao.org/evaluation/list/completed/en
GEF	https://www.gefio.org/
Green Climate Fund	https://ieu.greenclimate.fund/evaluations
IAEA	https://www.iaea.org/publications
IFAD	https://www.ifad.org/en/web/ioe
ILO	https://www.ilo.org/eval/Evaluationreports/lang--en/index.htm and i-eval Discovery (ilo.org)
IMO	https://www.imo.org/en/publications/Pages/Home.aspx
IOM	https://evaluation.iom.int
ITC	https://intracen.org/about-us/governance/evaluation
OCHA	https://www.unocha.org/evaluations-and-reviews
OHCHR	https://www.ohchr.org/en/about-us/evaluation-un-human-rights
OIOS	https://oios.un.org/inspection-evaluation-reports
Pan-American Health Organization (PAHO)	PAHO has an intranet-based platform of evaluation reports, currently for internal use.
UNAIDS	https://www.unaids.org/en/whoweare/evaluation
UNCDF	https://www.uncdf.org/evaluation
UNCTAD	https://unctad.org/about/evaluation
UNDP	https://aida.undp.org/landing
UN-DPA	https://dppa.un.org/en/planning-monitoring-and-evaluation
UN-DPKO	https://peacekeepingresourcehub.un.org/en/evaluation
UNECA (United Nations Economic Commission for Africa)	https://repository.uneca.org/
UNECE (United Nations Economic Commission for Europe)	https://unece.org/evaluation-reports
UN Evaluation Group (UNEG)	http://uneval.org/evaluation/reports
UNEP	https://wedocs.unep.org/handle/20.500.11822/1

UNESCO	https://www.unesco.org/en/ios/evaluation/reports?hub=67038
UNFPA	https://www.unfpa.org/evaluation/database
UN-Habitat	https://unhabitat.org/about-us/evaluation
UNHCR	https://www.unhcr.org/about-unhcr/who-we-are/evaluation-office
UNICEF	https://www.unicef.org/evaluation/reports#/
UNIDO	https://www.unido.org/resources/evaluation-and-internal-oversight/evaluation/resources
UNITAR	https://unitar.org/results-evidence-learning/evaluation
UNOCT	https://www.un.org/counterterrorism/publications
UNODC	https://www.unodc.org/unodc/en/evaluation/reports.html
UNRWA	https://dios.unrwa.org/evaluation-division
UN Women	https://gate.unwomen.org
WFP	https://www.wfp.org/publications
WHO	https://www.who.int/about/what-we-do/evaluation
WIPO	https://www.wipo.int/about-wipo/en/oversight/iaod/evaluation/
Note: The links to the following agencies could not found and will be updated when accessible: UN-DPI, UNV, CTBTO, ICAO, Office of Legal Affairs, OPCW, UN-DGACM, UNICRI, UN-PBSO, WMO, WTO, UNWTO, UNDESA, ICC	

Annex 2 List of databases for external research

Table x: Non exhaustive list of databases for research and non-UN evaluations

Database	Link
The Campbell Collaboration Library of Systematic Reviews	https://www.campbellcollaboration.org/better-evidence.html
3ie (International Initiative for Impact Evaluation) Development Evidence Portal	https://developmentevidence.3ieimpact.org/
AGRIS (International System for Agricultural Science and Technology)	https://www.fao.org/agris/
Web of Science	https://clarivate.com/webofsciencegroup/solutions/web-of-science
Medline (Via PubMed)	https://pubmed.ncbi.nlm.nih.gov/
EMBASE	http://www.elsevier.com/online-tools/embase
PsycINFO	http://www.apa.org/pubs/databases/psycinfo/index.aspx
Scopus	https://www.scopus.com/
USAID Evaluations Clearing house	http://dec.usaid.gov/
Organisation for Economic Co-operation and Development:	http://www.oecd.org/
Innovations for Poverty Action Publications:	https://www.poverty-action.org/publications
Environmental Evidence Library:	http://www.environmentalevidence.org/completedreviews