The Global SDG Synthesis Coalition

Social Assistance for the Furthest Behind

What Works and How Across SDGs 1-5

Appendices

Viviana Urueña, Atika Pasha, Johanna Kern, Zahra Kamal, Marina Rinke, Cristina Pozneanscaia

Social Assistance for the Furthest Behind —
What Works and How Across SDGs 1–5

Synthesis report of evaluative evidence related to the People Pillar of the Sustainable Development Goals.

Citation: Urueña V., Pasha A., Kern J., Kamal Z., Rinke M., & Pozneanscaia C. (2025). Social Assistance for the Furthest Behind: What Works and How Across SDGs 1–5. New York: The Global SDG Synthesis Coalition.

https://www.sdgsynthesiscoalition.org/pillar/people-pillar

Social Assistance for the Furthest Behind

What Works and How Across SDGs 1–5

Appendices

Viviana Urueña, Atika Pasha,
Johanna Kern, Zahra Kamal,
Marina Rinke,
Cristina Pozneanscaia

Contents

Appendix G:
Barriers & Drivers – UN evaluations 78

Appendix H:
System-level outcomes and program contributions 83

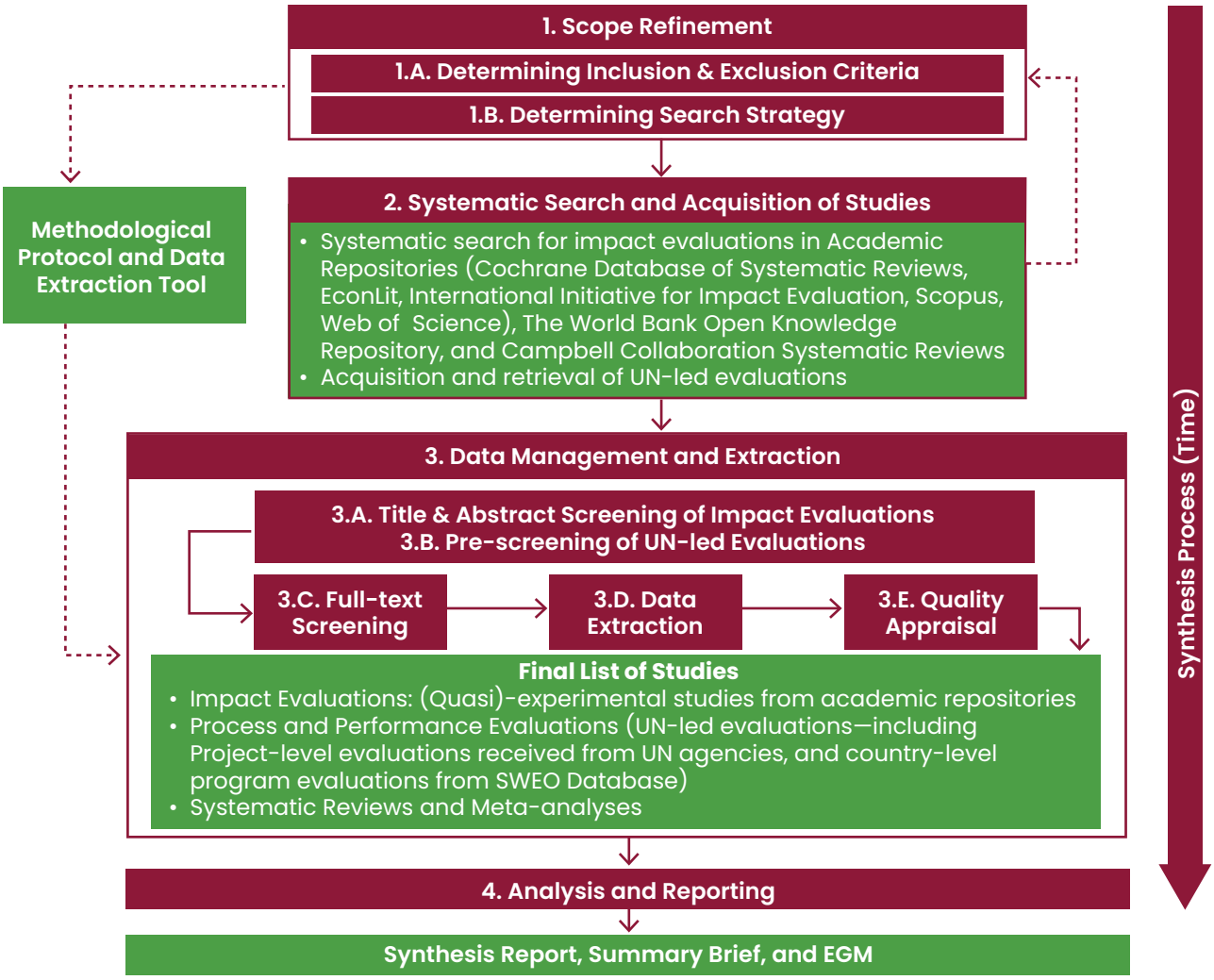
Appendix A: method

Methodological approach

The section below outlines the methodological approach used in this mixed-methods evidence synthesis and summarises the key stages of the process.

Figure 1.

Overview of synthesis methodology



Source: Synthesis team elaboration

Figure 1 illustrates the workflow for these stages, including the intermediate outputs generated at each step and the final products of the synthesis. The work proceeded through four main phases:

1. Defining the scope (October 2024–January 2025)
2. Systematic searching and study acquisition (February 2025)
3. Data management (February–July 2025)
4. Analysis and reporting (August–February 2026)

For impact evaluations and synthesis studies, an electronic systematic search in academic repositories (Cochrane Database of Systematic Reviews, EconLit, International Initiative for Impact Evaluation – 3ie, Scopus, Web of Science, and World Bank Open Knowledge Repository, and the Campbell Collaboration Systematic Reviews) was performed[1] For UN-led process and performance evaluations, a simple filter in the United Nations Sustainable Development Group System-Wide Evaluation Office (SWEO) Database was applied to identify relevant country-programme evaluations, while a list of project-level evaluations was provided directly by various UN agencies.[2]

All studies retrieved from the searched databases were imported into EPPI-Reviewer (Thomas et al. 2023), where all key data management steps were carried out. The processing of 4,202 impact evaluations, synthesis, and UN-led process and performance evaluations underwent five different stages throughout the data management and extraction phase. This five-step approach ensures that the findings are grounded in high-quality evidence within the social protection domain, providing valuable insights to policymakers, practitioners, and other stakeholders in large-scale social-protection programming. Each of these steps is explained in detail below.

The processing of the 4,202 studies underwent through five different steps throughout the data management and extraction phase:

i. Title and abstract (TiAb) screening of impact evaluations and syntheses: A two-stage pilot was completed before moving to single-screening. Priority screening feature was used in EPPI-Reviewer to rank and review the most relevant studies first, thereby optimising the efficiency of the review process. From 4,047 records identified after searching and deduplication, 1,085 studies (27%) were taken forward for full-text screening.

2. Pre-Screening of UN-led process and performance evaluations: UN-led evaluations underwent a separate pre-screening process. An automated text-screening step was used to refine the initial pool and identify evaluations likely to be relevant to the synthesis. This was followed by pilot title and executive summary screening and subsequent single screening. Additional automated refinement was conducted to align the sample with the planned scope. After applying relevance and quality criteria, 99 evaluations were retained for data extraction.

3. Full-text screening: A total of 1,085 impact evaluations and synthesis studies were assessed at full text, resulting in 242 initially eligible studies. To align with the intended scope and resource parameters, the time window for inclusion was narrowed to studies published from 2018 onwards, yielding a final set of 178 studies. During data extraction, a small number of additional ineligible studies were identified and removed, resulting in 154 eligible studies published between 2015 and 2024.

1 For details of the search strategy, please refer to the Methodology section in the protocol.

2 The United Nations Sustainable Development Group System-Wide Evaluation Office (SWEO) conducts system-wide evaluations and provides evidence on the contribution of the United Nations development system towards the achievement of the 2030 Agenda for Sustainable Development and the Sustainable Development Goals (SDGs).

4. Data Extraction: Data extraction protocols were informed by SPIDER and piloted separately for impact evaluations and syntheses and UN-led process and performance evaluations. Reviewers coded study characteristics, methods, implementation features, and findings relevant to the synthesis questions. During this stage, further refinements were made to ensure alignment with the phenomenon of interest and the SPIDER sampling criteria. The final dataset comprised 154 impact evaluations and synthesis studies and 94 UN-led evaluations.

5. Quality Appraisal: Quality appraisal was conducted in parallel with data extraction, following the procedures described in the protocol. Impact evaluations were assessed using a Risk of Bias (RoB) tool adapted for experimental and quasi-experimental designs. Studies were rated across key domains and assigned an overall RoB score. Synthesis studies were appraised using the 3ie SURE checklist, with reviews assigned High, Medium, or Low confidence ratings. UN-led process and performance evaluations relied on existing UN agency quality assessments, where available. Evaluations lacking an appraisal were assessed using a tool developed for the SDG 17 evidence synthesis. Only evaluations rated Highly Satisfactory or Satisfactory were included. The RoB and SURE quality appraisal tools can be found in Appendices B and C.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram (Appendix D) summarises the flow of studies through the screening process. Of 5,185 records identified, 4,202 remained after deduplication. Following title and abstract screening, and two rounds of text-screening for UN-led process and performance evaluations, 1,085 records proceeded to full-text review. After exclusions at this stage, 154 studies (106 impact studies and 48 syntheses) and 94 UN-led process and performance evaluations were included in the final evidence base.

Methodological limitations

This study is subject to several methodological limitations:

Not all studies within the included systematic reviews adhered to the SPIDER criteria. As a result, data extraction was limited to eligible studies only, leading to information gaps where systematic reviews did not report findings separately for those studies. Consequently, not all systematic reviews could be used in their entirety.

Identifying appropriate comparison groups proved difficult. In particular, it was not always straightforward to determine whether voucher-based or in-kind interventions were assessed against a no-intervention counterfactual or compared with alternative transfer modalities. While impacts were generally easier to interpret in the former case, interpretation was more complex in the latter, especially where interventions were implemented as top-ups to existing programmes. Despite rigorous quality assurance during data extraction, a small number of comparisons may not have been captured with full precision.

Applying an implementation-research lens to a body of existing process and performance evaluations limits the exploration of causal mechanisms, theory-based implementation pathways and system-level contribution. The heterogeneity of methods, inconsistent frameworks, and lack of uniform process-oriented designs in the underlying evaluations limited robust cross-study aggregation and synthesis. Implementation science literature highlights that such methodological and contextual variability reduces the certainty of inferences about how and why interventions work (see for example Nilsen 2015; Proctor et al. 2013).

The extent to which the UN-led process and performance evaluations cover relevant projects and factors for success varied considerably. Whereas evaluations of individual interventions (n=32) contained more relevant information, country programme, portfolio, and strategy evaluations (n=62) were more limited in this regard. In addition to interventions that are relevant for this study, they often examine a wide range of projects and themes that fall outside the scope. In many of the latter evaluations, information on relevant interventions represents only a small share of the overall content, or findings are presented in a highly aggregated manner, leading to difficulties in understanding whether and to what extent statements refer to relevant assistance protection interventions.

The breadth of the evidence base combined with practical constraints in time and resources established limitations to the iterative analysis and related to the rigour of qualitative analysis and coding across the evidence base. While the highest standards of qualitative research emphasise iterative analysis, i.e., revisiting the data multiple times to refine codes and ensure internal consistency, this level of analytic iteration was not fully feasible within the scope of this synthesis. The broad evidence base as well as time and resource constraints limited the extent to which evaluations could be reread and recoded as new insights emerged. Nonetheless, codes and categories were refined during subsequent stages of analysis to enhance consistency and coherence. Consequently, numerical counts or frequency references presented in this report should be understood as indicative of general trends and patterns rather than as precise quantitative measurements of the evidence base.

Despite these limitations, confidence in the findings remains Medium to High as the analysis draws primarily on robust impact studies and systematic reviews as well as process and performance assessed as having Satisfactory to High methodological quality. This synthesis also applies systematic screening, quality assurance, and transparent analytical procedures to mitigate risks to validity.

Appendix B: RoB tool

A. Study type	☐ Experimental ☐ Quasi-experimental – RDD ☐ Quasi-experimental – DiD ☐ Quasi-experimental – Matching ☐ Quasi-experimental – IV	
B. General questions	**1.** Are the mean values or the distributions of the covariates at baseline statistically different for the control or comparison group ($p<0.05$)?	☐ Yes ☐ No ☐ Unclear
	2. Are these differences controlled for using covariate analysis in the impact evaluation?	☐ Yes ☐ No ☐ Unclear
	3. Is difference-in-difference estimation used in addition to main specification?	☐ Yes ☐ No ☐ Unclear
	4. If the study does not use difference-in-difference, does it control for baseline values of the outcome of interest (ANCOVA)?	☐ Yes ☐ No ☐ Unclear

C. Attrition	**1.** Is the attrition rate from the study below 10%?	☐ Yes ☐ No ☐ Unclear
	2. Is the attrition rate statistically significantly different between the treatment and comparison groups?	☐ Yes ☐ No ☐ Unclear
D. Spillovers and Contamination	**1.** Are comparisons sufficiently isolated from the intervention (e.g., are the control or comparison groups sufficiently geographically separated)?	☐ Yes ☐ No ☐ Unclear
	2. Contamination: Does the control group receive the intervention?	☐ Yes ☐ No ☐ Unclear
	3. Contamination: If the control group receives the intervention but for a shorter amount of time, does the study assess the likelihood that the control group has received equal benefits as the treatment group?	☐ Yes ☐ No ☐ Unclear
E. Sample Size	**1.** Does the study account for lack of independence between observations within assignment clusters if the outcome variables are clustered?	☐ Yes ☐ No ☐ Unclear
	2. Is sample size likely to be sufficient to find significant effects of the intervention?	☐ Yes ☐ No ☐ Unclear

If "*Experimental*" is marked in set A,

F. Randomization	Does the study apply randomised assignment?	☐ Yes ☐ No ☐ Unclear

If "*Quasi-experimental – RDD*" is marked in set A,

G. RDD	Is the allocation of the programme based on a pre-determined continuity on a continuous variable and blinded to the beneficiaries, or, if not blinded, individuals cannot reasonably affect the assignment variable in response to knowledge of the participation rule?	☐ Yes ☐ No ☐ Unclear

If "*Quasi-experimental – DiD*" is marked in set A,

H. DiD	If the study is quasi-experimental and uses difference-in-difference estimation, is it showing that the parallel trends assumption is valid?	☐ Yes ☐ No ☐ Unclear

If "*Quasi-experimental – Matching*" is marked in set A,

I. Matching	Are the characteristics of the treatment and comparison group similar based on statistical significance tests after matching?	☐ Yes ☐ No ☐ Unclear

If "*Quasi-experimental – IV*" is marked in set A,

J. IV	**1.** Does the study describe clearly the instrumental variable(s) or identifier used and why it is exogenous?	☐ Yes ☐ No ☐ Unclear
	2. Are the instruments jointly significant at the level of F≥10? If an F test is not reported, does the author report and assess whether the R-squared of the instrumenting equation is large enough for appropriate identification (R-squared>0.5)?	☐ Yes ☐ No ☐ Unclear

Decision rule per bias for randomised studies

Risk of Selection Bias:	• **Low risk (1 pt):** Good RCT; RCT with moderate (10%) attrition but convincing robustness checks
	  • Questions B1 and B2 are replied to with "Yes" or B1 with "No"
	  • Question C1 is replied to with "Yes"
	  • Question C2 is replied to with "No"
	  • Question E1 is replied to with "Yes" or "Unclear"
	  • Question E2 is replied to with "Yes" or "Unclear"
	• **Medium risk (2 pt):** Underpowered RCT or with significant balance problems, moderate attrition, and no convincing robust checks
	  • Questions B1 and B2 are replied to with "Yes" or "Unclear"
	  • Question C1 is replied to with "Yes" - ONLY if attrition lower than 30%
	  • Question C2 is replied to with "No"
	  • Question E1 is replied to with "Yes" or "Unclear"
	  • Question E2 is replied to with "No"
	• **High risk (3 pt):** RCT with very large balance problems, high attrition, and no convincing robustness checks
	  • Questions B1 and B2 are replied to with "Unclear" or B1 is "Yes" and B2 with "No"
	  • Question C1 is replied to with "No"
	  • Question C2 is replied to with "Yes" or "Unclear"
	  • Question E1 is replied to with "No" or "Unclear"
	  • Question E2 is replied to with "No"

Risk of Performance Bias
 	
(related to contamination and spillovers)
	• **Low risk (1 pt):** Cluster-randomised controlled trial and no evidence that control group received the program. Researchers have a convincing methodology for estimating spillovers and no evidence that the control or comparison group received the program and beneficiaries are not in contact with non-beneficiaries.
	  • Question D1 is replied to with "Yes"
	  • Question D2 is replied to with "No"
	• **Medium risk (2 pt)**: If some percentage of the control or comparison group receives the intervention (<20%).
	  • Question D1 is replied to with "Yes" or "Unclear"
	  • Question D2 is replied to with "No" – ONLY if contamination is lower than 20%
	  • Question D3 is replied to with "Yes" or "Unclear"
	• **High risk (3 pt):** If a significant percentage (>20%) of the control or comparison group receives the intervention. If there is evidence that the treatment and control or comparison group compete for the same jobs (vocational training) or sales (business training).
	  • Question D1 is replied to with "No"
	  • Question D2 is replied to with "Yes"
	  • Question D3 is replied to with "Yes"

Risk of outcome and analysis reporting bias	• **Low risk (1 pt):** If there is no evidence for selective reporting • **Medium risk (2 pt):** if there is evidence that the authors prioritise outcome variables with statistically significant effects over outcome variables that do not show statistically significant effects but nonetheless report all effects in tables. • **High risk (3 pt):** If there is evidence that the authors prioritise outcome variables with statistically significant effects over outcome variables that do not show statistically significant effects and do not show the results of the non-statistically significant effects.
Risk of other bias	• **Low risk (1 pt):** If there is no evidence for other bias. • **Medium risk (2 pt):** if authors are supposed to cluster standard errors but fail to do so. • **High risk (3 pt):** If authors make other clear analytical mistakes.

Decision for aggregating scores

Risk of Bias	Criteria
Low risk	If total points <5
Medium	If total points <7
High risk	If total points >7

Decision rule per bias for quasi-experimental studies

Risk of Selection Bias	• **Low risk (Matching and Fuzzy RDDs are never low risk) – 1pt:**
	  • RDD *(only sharp RDD)*:
	  • Questions B1 and B2 are replied to with "Yes" or B1 with "No"
	  • Question C1 is replied to with "Yes"
	  • Question C2 is replied to with "No"
	  • Question E1 is replied to with "Yes" or "Unclear"
	  • Question E2 is replied to with "Yes" or "Unclear (sufficient sample size on both sides of the cut-off point)
	  • Question G is replied to with "Yes"
	  • if all the robustness checks support that the beneficiaries and comparison just above and just below the cut-off point are comparable to each other, there is a clear cut-off point
	  • IV:
	  • Questions B1 and B2 are replied to with "Yes" or B1 is "No"
	  • Question C1 is replied to with "Yes"
	  • Question C2 is replied to with "No"
	  • Question E1 is replied to with "Yes" or "Unclear"
	  • Question E2 is replied to with "Yes" or "Unclear
	  • ONLY if randomization is the instrumental variable

Risk of Selection Bias (cont.)	• DiD:
	  • Questions B1 and B2 are replied to with "Yes" or B1 is "No"
	  • Question C1 is replied to with "Yes"
	  • Question C2 is replied to with "No"
	  • Question E1 is replied to with "Yes" or "Unclear"
	  • Question E2 is replied to with "Yes" or "Unclear
	  • Question H is replied to with "Yes"
	• **Medium risk (2pt):**
	  • RDD *(sharp or fuzzy)*:
	  • Questions B1 and B2 are replied to with "Yes" or "Unclear"
	  • Question C1 is replied to with "Yes" - ONLY if attrition lower than 30%
	  • Question C2 is replied to with "No"
	  • Question E1 is replied to with "Yes" or "Unclear"
	  • Question E2 is replied to with "No" - RDD with a lower sample on both sides of the cut-off point.
	  • Question G is replied to with "Yes" or "Unclear"
	  • IV
	  • Questions B1 and B2 are replied to with "Yes" or "Unclear"
	  • Question C1 is replied to with "Yes" - ONLY if attrition lower than 30%
	  • Question C2 is replied to with "No"
	  • Question E1 is replied to with "Yes" or "Unclear"
	  • Question E2 is replied to with "No"
	  • Question J1 is replied to with "Yes" or "Unclear"
	  • Question J2 is replied to with "Yes" or "Unclear"
	  • DiD:
	  • Questions B1 and B2 are replied to with "Yes" or "Unclear"
	  • Question C1 is replied to with "Yes" - ONLY if attrition lower than 30%
	  • Question C2 is replied to with "No"
	  • Question E1 is replied to with "Yes" or "Unclear"
	  • Question E2 is replied to with "No"
	  • Question H is replied to with "No"
	  • Controls for trends before the start of the program
	  • Matching:
	  • Questions B1 and B2 are replied to with "Yes" or "Unclear"
	  • Question C1 is replied to with "Yes" - ONLY if attrition lower than 30%
	  • Question C2 is replied to with "No"
	  • Question E1 is replied to with "Yes" or "Unclear"
	  • Question E2 is replied to with "No"
	  • Question I is replied to with "Yes"
	  • Matching algorithm uses baseline variables

Risk of Selection Bias (cont.)	• **High risk (3 pt):**
	  • RDD
	  • Questions B1 and B2 are replied to with "Unclear"
	  • Question C1 is replied to with "No"
	  • Question C2 is replied to with "Yes" or "Unclear"
	  • Question E1 is replied to with "No" or "Unclear"
	  • Question E2 is replied to with "No"
	  • Question G is replied to with "No"
	  • IV
	  • Questions B1 and B2 are replied to with "Unclear"
	  • Question C1 is replied to with "No"
	  • Question C2 is replied to with "Yes" or "Unclear"
	  • Question E1 is replied to with "No" or "Unclear"
	  • Question E2 is replied to with "No"
	  • Question J1 is replied to with "No"
	  • Question J2 is replied to with "No"
	  • DiD:
	  • Questions B1 and B2 are replied to with "Unclear"
	  • Question C1 is replied to with "No"
	  • Question C2 is replied to with "Yes" or "Unclear"
	  • Question E1 is replied to with "No" or "Unclear"
	  • Question E2 is replied to with "No"
	  • Question H is replied to with "No"
	  • Study does not control for trends before the start of the program
	  • Matching:
	  • Questions B1 and B2 are replied to with "Unclear"
	  • Question C1 is replied to with "No"
	  • Question C2 is replied to with "Yes" or "Unclear"
	  • Question E1 is replied to with "No" or "Unclear"
	  • Question E2 is replied to with "No"
	  • Question I is replied to with "Yes"
	  • Matching algorithm does not use baseline variables

Risk of Performance Bias
 	
(related to contamination and spillovers)
	• **Low risk (1 pt):**
	  • Question D1 is replied to with "Yes"
	  • Question D2 is replied to with "No"
	  • Questions D2 & D3 is replied to with "Yes" BUT researchers have a convincing methodology for estimating spillovers
	• **Medium risk (2 pt)**: If some percentage of the control or comparison group receives the intervention (<20%).
	  • Question D1 is replied to with "Yes" or "Unclear"
	  • Question D2 is replied to with "No" – ONLY if contamination <20%
	  • Question D3 is replied to with "Yes" or "Unclear"
	• **High risk (3 pt):** If a significant percentage (>20%) of the control or comparison group receives the intervention. If there is evidence that the treatment and control or comparison group compete for the same jobs (vocational training) or sales (business training).
	  • Question D1 is replied to with "No"
	  • Question D2 is replied to with "Yes"
	  • Question D3 is replied to with "Yes"

Risk of Outcome and Analysis Reporting Bias	• **Low risk (1 pt):** if there is no evidence for selective reporting • **Medium risk (2 pt):** if there is evidence that the authors prioritise outcome variables with statistically significant effects over outcome variables that do not show statistically significant effects but nonetheless report all effects in tables • **High risk (3 pt):** If there is evidence that the authors prioritise outcome variables with statistically significant effects over outcome variables that do not show statistically significant effects and do not show the results of the non-statistically significant effects
Risk of Other Bias	• **Low risk (1 pt):** If there is no evidence for other bias • **Medium risk (2 pt):** If the authors are supposed to cluster standard errors but fail to do so. • **High risk (3 pt):** If the authors make other clear analytical mistakes

Decision for aggregating scores

Risk of Bias	Criteria
Low risk	If total points <5
Medium risk	If total points <7
High risk	If total points >7

Appendix C: SURE tool

Section A: Methods used to identify, include, and critically appraise studies

A.1 Were the criteria used for deciding which studies to include in the review reported? Did the authors specify: ☐ Types of studies ☐ Participants/ settings/ population ☐ Intervention(s) ☐ Outcome(s)	☐ Yes ☐ Partially ☐ No **Coding guide - check the answers above** YES: All four conditions should be yes NO: All four conditions should be no PARTIALLY: All other cases
Comments (note important limitations or uncertainty)	
A.2 Was the search for evidence reasonably comprehensive? Were the following done: ☐ Language bias avoided (no restriction of inclusion based on language) ☐ No restriction of inclusion based on publication status ☐ Relevant databases searched (Minimum criteria: All reviews should search at least one source of grey literature such as Google; for health: Medline/Pubmed + Cochrane Library; for social sciences IDEAS + at least one database of general social science literature and one subject specific database) ☐ Reference lists in included articles checked ☐ Authors/experts contacted	☐ Yes ☐ Partially ☐ No ☐ Can't tell **Coding guide - check the answers above** YES: All five conditions should be yes PARTIALLY: Only relevant databases and reference lists are both reported, while other conditions not met NO: All other cases CAN'T TELL: If unclear
Comments (note important limitations or uncertainty)	

A.3 Does the review cover an appropriate time period?

- ☐ Is the search period comprehensive enough that relevant literature is unlikely to be omitted?

- ☐ Yes
- ☐ Can't tell (only use if no information about time period for search)
- ☐ No
- ☐ Unsure

Coding guide:

YES: Literature search goes back at least 15 years from SR's publication date

NO: Literature search does not go back 15 years from SR's publication date

CAN'T TELL: No information about time period for search

Report the time period for the search in the comment box.

Comments (note search period, any justification provided for the search period, or uncertainty)

A.4 Was bias in the selection of articles avoided?

Did the authors specify:

- ☐ Independent screening of full text by at least two reviewers
- ☐ List of included studies provided

- ☐ Yes
- ☐ No

Coding guide:

YES: Both conditions should be yes

NO: If only one or zero conditions are yes

Comments (note important limitations or uncertainty)

A.5 Did the authors use appropriate criteria to assess the quality and risk of bias in analysing the studies that are included?

- ☐ The criteria used for assessing the quality/risk of bias were reported
- ☐ A table or summary of the assessment of each included study for each criterion was reported
- ☐ Sensible criteria were used that focus on the quality/ risk of bias (and not other qualities of the studies, such as precision or applicability/external validity). "Sensible" is defined as a recognised quality appraisal tool/checklist, or similar tool which assesses bias in included studies. Please see note below for details of the main types of bias such a tool should assess.

- ☐ Yes
- ☐ Partially
- ☐ No

Coding guide:

YES: All three conditions should be yes

PARTIALLY: The first and third condition should be met. If the authors report the criteria for assessing risk of bias and report a summary of this assessment for each criterion, but the criteria may be only partially sensible (e.g., do not address all possible risks of bias, but do address some), downgrade to PARTIALLY.

NO: All other cases

Comments (note important limitations or uncertainty)

A.6 Overall, how much confidence do you have in the methods used to identify, include, and critically appraise studies?

Summary assessment score A relates to the 5 questions above.

- High confidence applicable when the answers to the questions in section A are all assessed as YES
- Low confidence applicable when any of the following are assessed as NO above: not reporting explicit selection criteria (A1), not conducting reasonably comprehensive search (A2), not avoiding bias in selection of articles (A4), not assessing the risk of bias in included studies (A5)
- Medium confidence applicable for any other – i.e., section A3 is assessed as NO or CAN'T TELL and remaining sections are assessed as 'YES, PARTIALLY, or CAN'T TELL

☐ **Low confidence** (limitations are important enough that the results of the review are not reliable)

☐ **Medium confidence** (limitations are important enough that it would be worthwhile to search for another systematic review and to interpret the results of this review cautiously if a better review cannot be found

☐ **High confidence** (only minor limitations)

Comments (note important limitations):

Note: Risk of bias is the extent to which bias may be responsible for the findings of a study. Bias is a systematic error or deviation from the truth in results or inferences. In studies of the effects of social, economic, and healthcare interventions, the main types of bias arise from systematic differences in the groups compared (selection bias), the intervention provided, or exposure to other factors apart from the intervention of interest (performance bias/contamination), withdrawals or exclusions of people entered into a study (attrition bias), or how outcomes are assessed (detection bias) and reported (reporting bias). Reviews of social science studies may be particularly affected by reporting bias, where a biased subset of all the relevant data and analyses is presented. Assessments of the risk of bias are sometimes also referred to as assessments of the validity or quality of a study. Validity is the extent to which a result (of a measurement or study) is likely to be true. Quality is a general notion of the strength or validity of a study, often indicating the extent of control over bias.

Section B: Methods used to analyse the findings

B.1 Were the characteristics and results of the included studies reliably reported?

Was there:

- ☐ Independent data extraction by at least two reviewers?
- ☐ A table or summary of the characteristics of the participants, interventions, and outcomes for the included studies?
- ☐ A table or summary of the results of all the included studies?

- ☐ Yes
- ☐ No
- ☐ Partially
- ☐ Not applicable (e.g., no included studies)

Coding guide:

YES: All three conditions should be YES

PARTIALLY: Conditions one and three are YES, but some information is lacking on second condition.

No: None of these are reported, or condition one is not met.

NOT APPLICABLE: if no studies/no data

Comments (note important limitations or uncertainty):

B.2 Are the methods used by the review authors to analyse the findings of the included studies clear, including methods for calculating effect sizes if applicable?

- ☐ Yes
- ☐ Partially
- ☐ No
- ☐ Not applicable (e.g., no studies or no data)

Coding guide:

YES: Methods used clearly reported. If it is clear that the authors use narrative synthesis, they do not need to say this explicitly.

PARTIALLY: Some reporting on methods but lack of clarity

NO: Nothing reported on methods

NOT APPLICABLE: if no studies/no data

Comments (note important limitations or uncertainty):

B.3 Did the review describe the extent of heterogeneity?

- ☐ Did the review ensure that included studies were similar enough that it made sense to combine them, sensibly divide the included studies into homogeneous groups, or sensibly conclude that it did not make sense to combine or group the included studies?

- ☐ Yes
- ☐ Partially
- ☐ No
- ☐ Not applicable (e.g., no studies or no data)

☐ Did the review discuss the extent to which there were important differences in the results of the included studies?

If a meta-analysis was done, was the I^2 chi square test for heterogeneity or other appropriate statistic reported? If no statistical test was reported, is a qualitative justification made for the use of random effects?

Coding guide:

YES: First two conditions should be YES, and third condition should be YES if applicable

PARTIALLY: The first category is yes and all other categories are NO or PARTIAL

NO: All other cases

☐ NOT APPLICABLE: if no studies/no data

Comments (note important limitations or uncertainty)

B.4 Were the findings of the relevant studies combined (or not combined) appropriately relative to the primary question the review addresses and the available data?

How was the **data analysis** done?

☐ Descriptive only

☐ Vote counting based on direction of effect

☐ Vote counting based on statistical significance

☐ Description of range of effect sizes

☐ Meta-analysis

☐ Meta-regression

☐ Other: specify

☐ Not applicable (e.g., no studies or no data)

How were the studies **weighted** in the analysis?

☐ Equal weights (this is what is done when vote counting is used)

☐ By quality or study design (this is rarely done)

☐ Inverse variance (this is what is typically done in a meta-analysis)

☐ Number of participants (sample size)

☐ Other: specify

☐ Not clear

☐ Not applicable (e.g., no studies or no data)

☐ Yes

☐ Partially

☐ No

☐ Not applicable (e.g., no studies or no data)

☐ Can't tell

Coding guide:

YES: If appropriate table, graph or meta-analysis AND appropriate weights AND extent of heterogeneity taken into account

NO: If narrative OR vote counting (where quantitative analyses would have been possible) OR inappropriate table, graph or meta-analyses OR unit of analysis errors not addressed (and should have been).

NOT APPLICABLE: if no studies/no data

PARTIALLY/CAN'T TELL: if unsure (note reasons in comments below)

Did the review address **unit of analysis errors?**

- ☐ Yes - took clustering into account in the analysis (e.g., used intra-cluster correlation coefficient)
- ☐ No, but acknowledged problem of unit of analysis errors
- ☐ No mention of issue

Not applicable - no clustered trials or studies included

Comments (note important limitations or uncertainty)

B.5 Does the review report evidence appropriately?

- ☐ The review makes clear which evidence is subject to low risk of bias in assessing causality (attribution of outcomes to intervention) and which is likely to be biased and does so appropriately
- ☐ Where studies of differing risk of bias are included, results are reported and analysed separately by risk of bias status

- ☐ Yes
- ☐ No
- ☐ Partially
- ☐ Not applicable

Coding guide:

YES: Both conditions fulfilled, or only first condition fulfilled if no studies with different risk of bias

NO: No conditions fulfilled

PARTIALLY: Only one condition fulfilled even though both are applicable, or when quality of study reporting limited

- ☐ NOT APPLICABLE: No included studies

Comments (note important limitations or uncertainty)

Please specify included study designs and any other comments (note important limitations or uncertainty):

B.6 Did the review examine the extent to which specific factors might explain differences in the results of the included studies?

- ☐ Were factors the review authors considered likely explanatory factors clearly described?
- ☐ Was a sensible method used to explore the extent to which key factors explained heterogeneity?
  - ☐ Descriptive/textual
  - ☐ Graphical
  - ☐ Meta-analysis by sub-groups
  - ☐ Meta-regression
  - ☐ Other

- ☐ Yes
- ☐ Partially
- ☐ No
- ☐ Not applicable

Coding guide:

YES: Explanatory factors clearly described and the methods used explore heterogeneity

PARTIALLY: Explanatory factors described but for meta-analyses, sub-group analysis or meta-regression not considered (when they should have been)

NO: No description or analysis of likely explanatory factors

	NOT APPLICABLE: e.g., too few studies, no important differences in the results of the included studies, or the included studies were so dissimilar that it would not make sense to explore heterogeneity of the results
Comments (note important limitations or uncertainty):	
B.7 Overall - how much confidence do you have in the methods used to analyse the findings relative to the primary question addressed in the review? Summary assessment score B relates to the six questions in this section, regarding the analysis. ☐ *High confidence* applicable *when all the answers to the questions in section B are assessed as YES* ☐ *Low confidence* applicable *when any of the following are assessed as NO above: critical characteristics of the included studies not reported (B1), not describing the extent of heterogeneity (B3), combining results inappropriately (B4), or reporting evidence inappropriately (B5)* • *Medium confidence applicable for any other*	☐ **Low confidence** (limitations are important enough for the results of the review not to be reliable) ☐ **Medium confidence** (limitations are important enough that it would be worthwhile to search for another systematic review and to interpret the results of this review cautiously if a better review cannot be found) ☐ **High confidence** (only minor limitations)
Comments (note important limitations or uncertainty)	
Use comments to specify if relevant, to flag uncertainty or need for discussion	

Section C: Overall assessment of the reliability of the review

C.1 Are there any other aspects of the review not mentioned before that lead you to question the results?	☐ One person performing title and abstract screening ☐ Robustness ☐ Interpretation ☐ Conflicts of interest (of the review authors or for included studies)
C.2 Are there any mitigating factors which should be taken into account in determining the reviews reliability?	☐ Limitations acknowledged ☐ No strong policy conclusions drawn (including in abstract/ summary)

Use comments to specify if relevant, to flag uncertainty or need for discussion

C.3 Based on the above assessments of the methods, how would you rate the reliability of the review?

☐ **Low confidence in conclusions about effects**:

The systematic review has the following major limitations…

☐ **Medium confidence in conclusions about effects**:

The systematic review has the following limitations…

☐ **High confidence in conclusions about effects**:

If applicable: The review has the following minor limitations…

Coding guide:

High confidence in conclusions about effects: High confidence noted overall for sections A and B, unless moderated by answer to C1

Medium confidence in conclusions about effects: Medium confidence noted overall for sections A or B, unless moderated by answer to C1 or C2

Low confidence in conclusions about effects: Low confidence noted overall for sections A or B, unless moderated by answer to C2

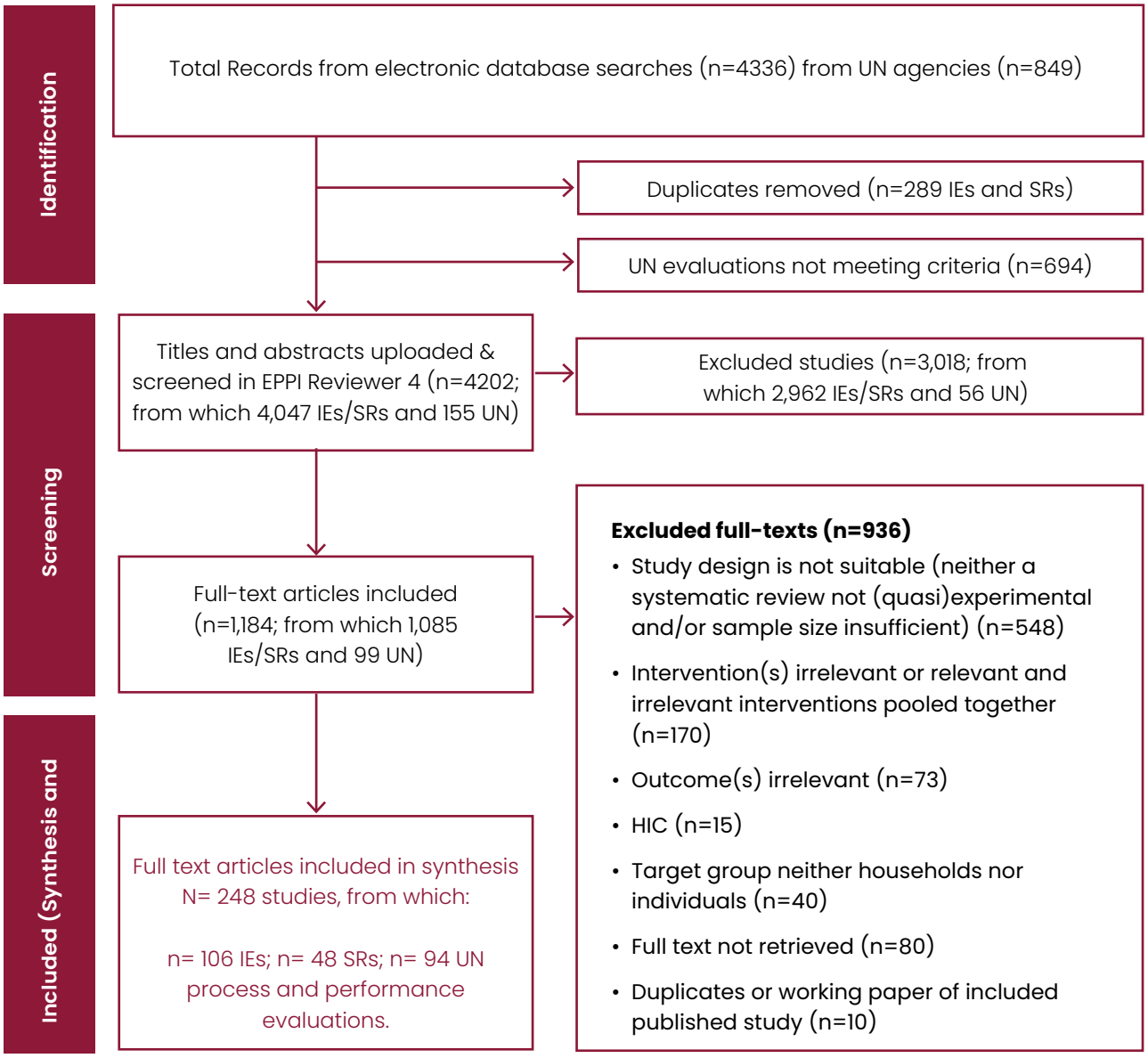
Where Moderated means that if any condition in C1 is met, the SR is rated one confidence level lower (e.g., from Medium to Low). If all conditions in C2 are met, the SR is rated one confidence level higher (e.g., from Medium to High).

Limitations should be summarised above, based on what was noted in Sections A, B and C.

Appendix D: PRISMA

Figure 2.

PRISMA Diagram



From Page et al., (2021). For more information, visit: http://www.prisma-statement.org

Sources: Synthesis team elaboration

Appendix E: Characteristics of included evaluations and studies

Figure 3 shows the distribution of the evidence base used in this synthesis across the years 2015-2024, disaggregated by study type. Accordingly, the body of evidence on gender-, age-, and disability-responsive voucher and in-kind transfer interventions appears to have grown steadily over the past decade, with a sharp increase in 2022 followed by a notable decline in 2024.[3] Notably, the number of UN-led evaluations in this area increased significantly—from only one in 2015 to a peak of 31 in 2023.

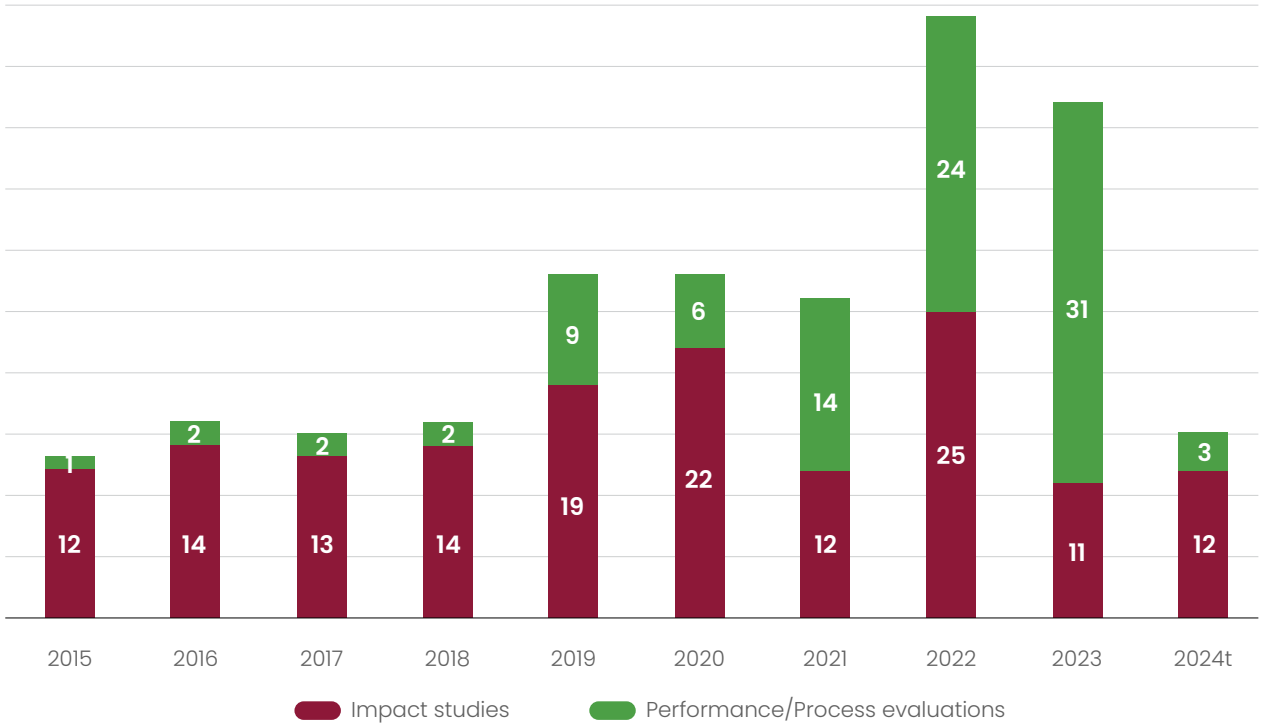
To contextualise the distribution of UN-led process and performance evaluations included in the synthesis across agencies, it is important to highlight that half the evaluations are from WFP and 13% from UNICEF. The synthesis also includes evaluations from UNFPA and FAO (5 each), UNDP (4), IFAD (3), ILO (2), with the remainder constituting a mix of other UN agency evaluations. Only three studies were commissioned by more than one UN body. The diversity of agencies reflects the broad engagement of UN entities in implementing and assessing voucher and in-kind transfer programmes.

UN-led process and performance evaluations typically take a holistic approach, with most studies included in the synthesis addressing mostly four of the six OECD DAC evaluation criteria. More specifically, 100% cover effectiveness, 95% address relevance and sustainability, and 87% examine coherence. The largest body of evidence relates to the conditions (barriers and drivers) influencing effectiveness, suggesting that donors and implementers place the greatest emphasis on measuring and understanding outcomes related to this particular dimension above all others.

3 The decline observed in 2024 does not necessarily reflect a reduction in programme implementation or evaluation efforts; rather, it may be due to incomplete database coverage of studies published in 2024, particularly as the searches were conducted in early 2025.

Figure 3.

Evidence base by type and publication year (n=248)



Source: Synthesis team illustration

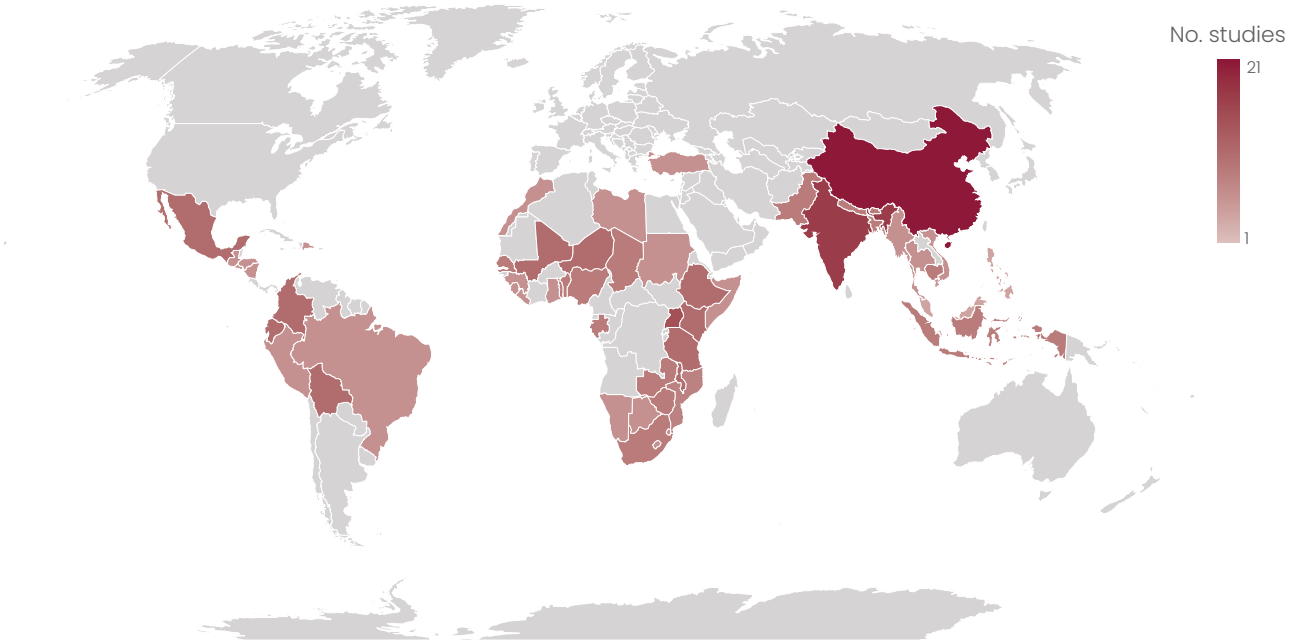
Geographic distribution

Similar to the findings from other syntheses efforts on social protection interventions (see e.g., Pasha, et al. 2023), most evidence is concentrated in Sub-Saharan Africa (119 out of 248). This is followed by South Asia (68), Latin America and the Caribbean (58), East Asia and the Pacific (55), the Middle East and North Africa (19), and Europe and Central Asia (11).[4] In Sub-Saharan Africa, evidence is relatively evenly distributed across three countries: Uganda (23), Kenya (18), and Burkina Faso (16). In South Asia, India stands out as the primary contributor, with 30 evaluations, followed by 15 in Nepal. In Latin America and the Caribbean, Mexico has the largest number of evaluations (18), followed by Colombia (13) and Ecuador (9). In East Asia and the Pacific, China (30) dominates the evidence base. Notably, impact studies are predominantly concentrated in China and India, while UN-led process and performance evaluations are more geographically dispersed, with comparatively less emphasis on these two countries.

4 Regional totals do not add up to 248 as the evidence base contains systematic reviews covering multiple regions and multi-country studies that span more than one region.

Figure 4.

Geographic distribution of evidence (n=200)



Source: Synthesis team illustration. Map does not include SRs as these usually include studies from multiple countries.

As seen in Figure 4, the majority of the evidence (39% of all studies) is concentrated in lower-middle-income countries. Following this, low-income and upper-middle-income countries each have similar shares of total studies (23%), but the former have a slightly higher proportion of UN-led evaluations, while the latter have a slightly higher proportion of impact studies. The remaining 15% cover a mix of income categories.

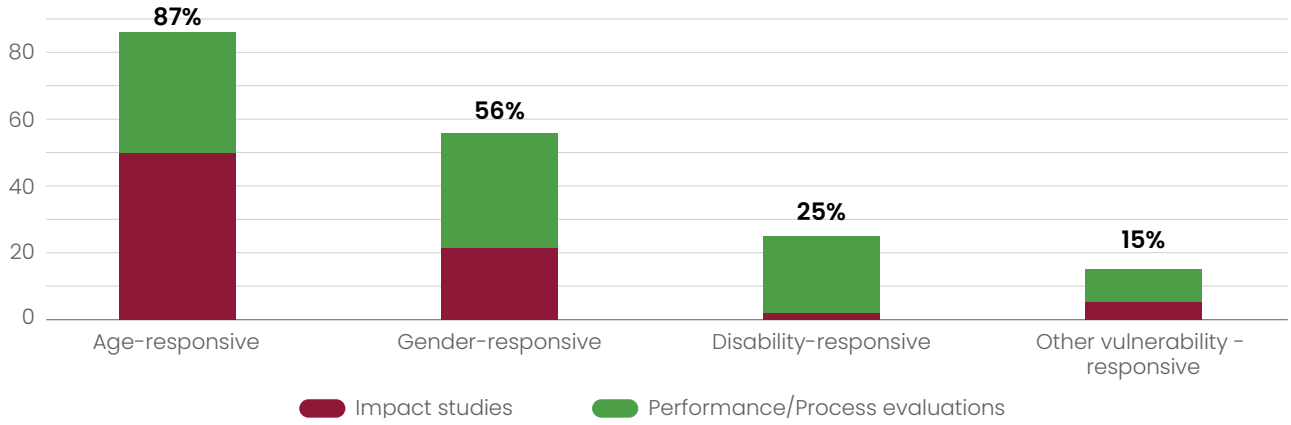
Despite their heightened vulnerability and greater need for effective social protection, the lower share of rigorous evidence from low-income countries (23%) points to a programming or research gap. This limited coverage may be partly attributable to the fragility of many low-income settings, where conflict and weak institutional capacity often hinder data collection and evaluation efforts or the establishment of a social assistance system that delivers vouchers or in-kind transfers. For instance, 64% of the studies included in the synthesis focus exclusively on countries without institutional fragility or conflict, 23% examine countries affected by violent conflict, 3% focus on those experiencing high levels of institutional fragility, and 9% address contexts characterised by both. Regarding the location within countries, most evidence is collected from rural areas (41%), 34% of studies cover both rural and urban contexts, and 6% are predominantly urban. For 16% of the studies, this information is not reported.

Target groups

The majority of interventions can be classified as age-responsive (87%), with children (0–13 years) identified as the primary age group among targeted beneficiaries, followed by adolescents (14–17 years). Figure 5 shows that over half of the interventions are identified as gender-responsive (56%), with a particular focus on benefiting women, including mothers and caregivers. In contrast, disability-responsive interventions are less prevalent, representing only 25% of the evidence base. An additional 15% of interventions target other vulnerable sub-populations, such as those living in extreme poverty or those affected by humanitarian crises

(including refugees and internally displaced persons), rural and remote populations, or individuals identified as sex workers. These findings highlight the prioritization of children and adolescents in social protection programming, reflecting an alignment with SDGs. **However, there is a relative underrepresentation of interventions specifically designed for elderly (65+) and persons with disabilities, two groups often facing heightened vulnerability.**

Figure 5.



Evidence by eligibility criteria (n=248 studies)

Source: Synthesis team illustration. It is important to note that 49% of studies applied multiple eligibility criteria. This justifies the decision to present overlapping categories and explains why percentages do not add up to 100%.

Compared to impact studies, UN-led process and performance evaluations (and likely UN programming) are more likely to target highly vulnerable populations, including disability-responsive programmes and those designed for other marginalised or at-risk groups. **However, emerging findings from the qualitative analysis of these evaluations suggest that responsiveness of interventions towards vulnerable groups was often stated as an objective but not reported at implementation or outcome level.** This points to a high commitment to the principle of Leaving No One Behind within UN programming, but also highlights persistent evidence gaps on effective implementation and disaggregated outcome reporting.

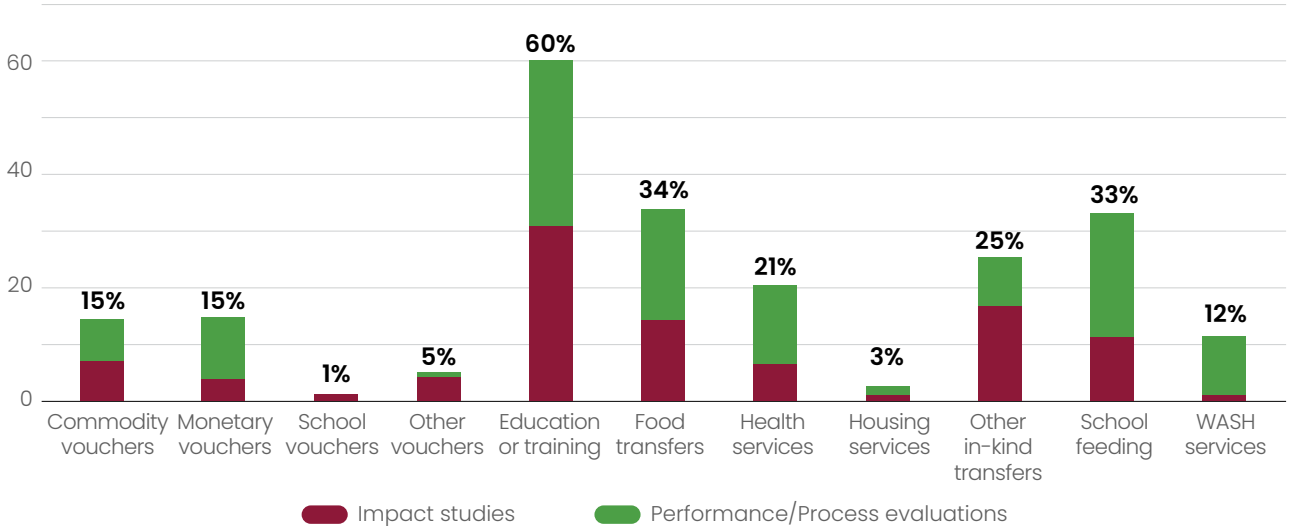
Interventions and outcomes

In-kind transfer interventions appear more frequently in the evidence base (228 studies) compared to voucher interventions (77), with education and training (across SDGs) making up the largest share (60%) (Figure 6). For voucher interventions, the evidence shows no difference in the proportion of interventions studying monetary or commodity vouchers (37 and 36 studies, respectively). Compared to the impact studies, water, sanitation and hygiene (WASH), health services, and monetary voucher interventions appeared more frequently in the performance and process evaluations.[5]

5 A monetary voucher is a voucher that has a fixed monetary value and can be exchanged for goods or services up to that value, typically at authorized vendors, but cannot be redeemed for cash.

Figure 6.

Evidence by intervention type (n=248 studies)



Source: Synthesis team illustration. Note that the majority of studies included multiple intervention types. This justifies our decision to present overlapping categories and explains why percentages do not add up to 100%.

The distribution of key implementing actors shows a diverse and multi-stakeholder ecosystem. Governmental bodies (local or national) are identified as key actors in 53% of total studies, with the largest share involving collaborations with non-governmental partners (33% of total studies). Such collaborations are particularly prevalent in UN-led process and performance evaluations. This finding reflects the central role of public systems, often in partnership with UN agencies, in delivering social assistance programmes. Standalone government-led programmes are present in 12% of the evidence base and are more often found in the impact studies (19%) than in UN-led process and performance evaluations included in this synthesis (2%). Researchers are identified as key implementing actors in 17% of total studies (27% of the impact studies, particularly in those with experimental designs). A multi-stakeholder model with all three actors is less prevalent but present in 7% of the studies.

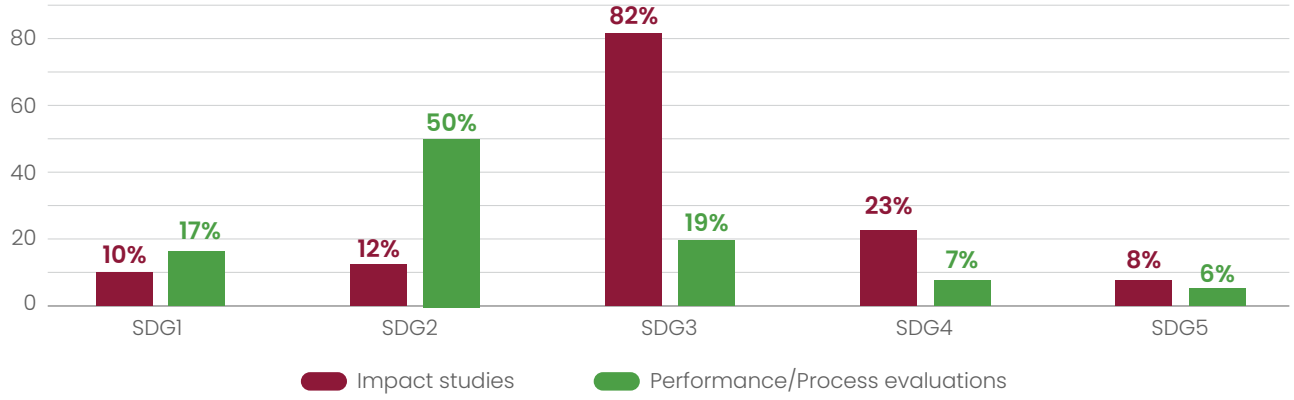
Overall, governments, NGOs, and researchers each play a critical role in advancing the SDGs, underscoring the importance of multi-stakeholder collaboration in scaling and evaluating inclusive interventions that are age-, gender-, and disability-responsive, with the aim of Leaving No One Behind.

In terms of outcome categories, evidence is strongly concentrated around the effectiveness of health (SDG 3) and zero hunger (SDG 2) (see Figure 7). The evidence is concentrated around SDG 3, with related outcomes found in 82% of impact studies and in 19% of UN-led process and performance evaluations. Within SDG 3, child health subcategory is the most frequently found, followed by nutrition and sexual and reproductive health. SDG 2 follows and is found in 12% of impact studies and in 50% of UN-led process and performance evaluations. Most studies examine food consumption-related outcomes. SDG 4 is more strongly represented in the impact studies (23%), with most studies focusing on learning outcomes and achievement, particularly through standardised test scores. In contrast, SDG 5 is found in 8% of impact studies and 6% of process and performance evaluations. Evaluations under this category primarily examine outcomes related to intimate partner violence. These findings align with other syntheses (see, for example, Pasha et al. 2023 for an EGM on cash transfers), where the health category is the most extensively examined and women's empowerment among the least.

The majority of UN-led process and performance evaluations (83%) assesses effects on (social protection) system-level changes.[6] Most intervention efforts and their evaluations (47%) concentrate on improving the implementation of existing laws and policies (including service delivery and budgeting), closely followed by efforts to develop new policies (42%). Efforts to establish or change laws are relatively less prominent (16%). About one third (31%) of programmes and their evaluations target more than one level of system change.

Figure 7.

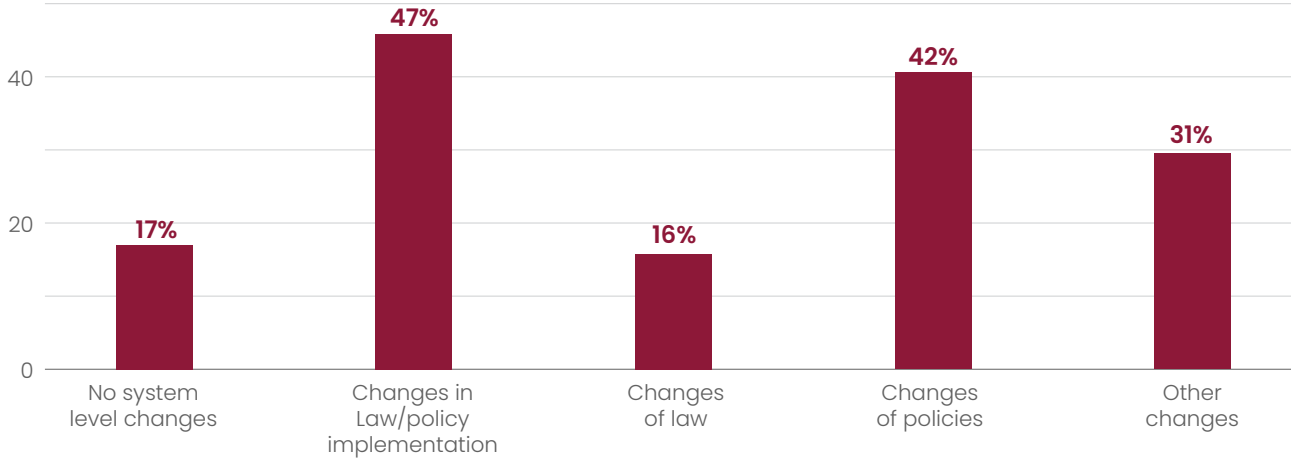
Evidence by outcome type (n=248)



Source: Synthesis team illustration Many studies examine multiple outcomes, which explains why percentages add up to >100%.

Figure 8.

Evidence by system-level change outcomes (n=94)



Source: Synthesis team illustration Many studies examine multiple outcomes, which explains why percentages add up to >100%.

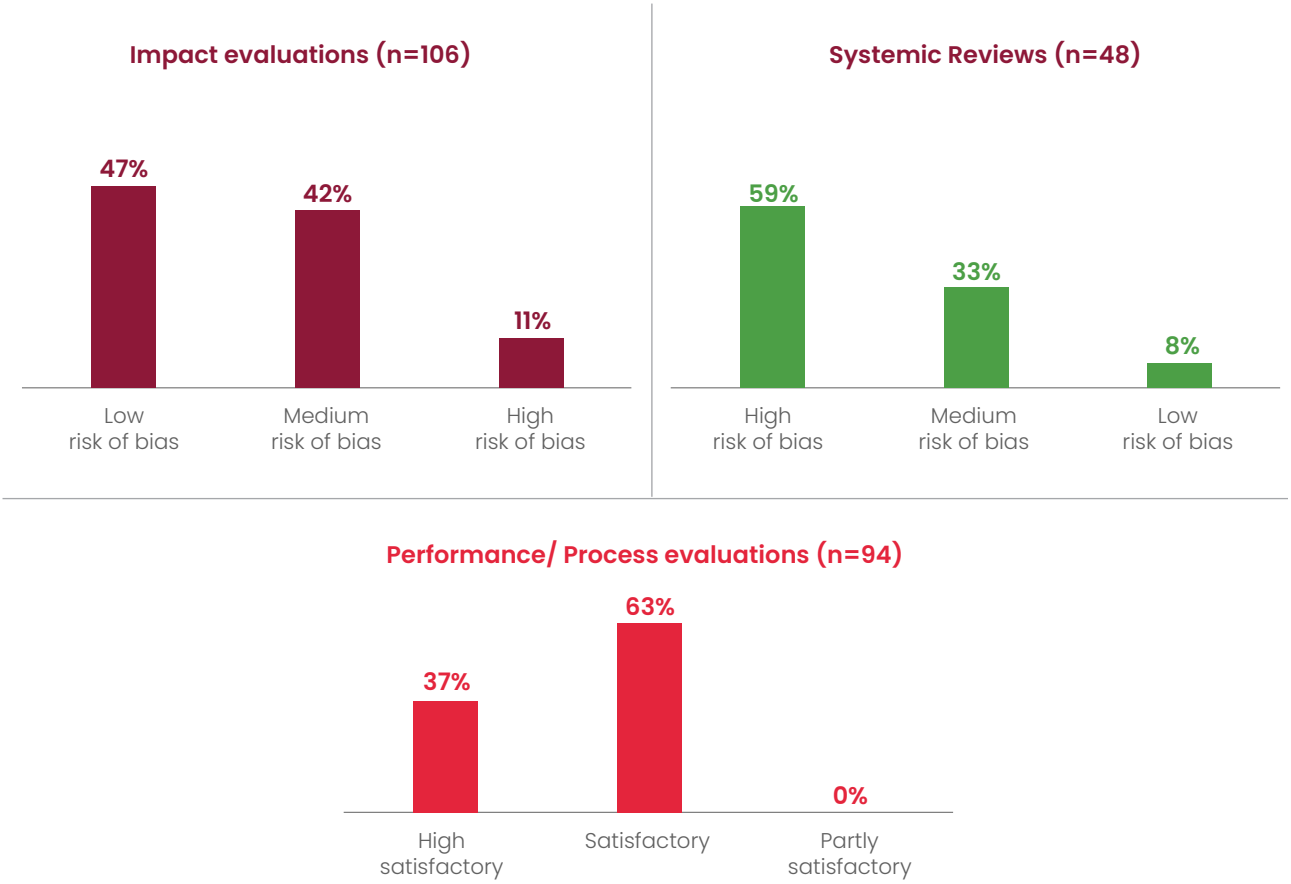
6 The synthesis does not examine beneficiary-level outcomes categorized along SDG 1-5 of UN-led evaluations as the added value and thereby focus of the qualitative analysis of UN-led evaluations was on assessing implementation science as well as system-level changes.

Quality of studies

The overall confidence in the findings of this synthesis is rated as Medium to High. Only a small share of impact evaluations was assessed as having risk of bias (RoB), mainly due to attrition or contamination, indicating a rigorous quality appraisal process. Confidence is further strengthened by the predominance of experimental designs (70 experimental versus 36 quasi-experimental) and by the high confidence ratings assigned to most synthesis reviews (using the 3ie SURE checklist). In addition, only UN-led process and performance evaluations rated Satisfactory or above were included. Taken together, this evidence base provides a strong foundation for drawing reliable conclusions and informing policy and programming decisions.

Figure 9.

Quality Assurance



Source: Synthesis team illustration

Appendix F: SDG effectiveness tables

Summary of gender-responsive interventions – SDG 3

	Child health	SRH and maternal health	Mental health and well-being	Access and use of health services	Nutrition and dietary diversity	Child labour	Other health outcomes
# of studies	18	22	4	2	10	2	3
% of outcomes measured are effective	44%	49%	29%	40%	46%	-	-
	Evidence positive & weak	Evidence positive & strong for vouchers but positive & weak for in-kind	Evidence positive & weak	Evidence positive & weak	Evidence positive & weak	Mixed results	Mixed results

Source: Synthesis team elaboration

Notes: Positive & strong: >50% outcomes are positive and significant; Positive & weak: <50% outcomes but >than 30%; Not effective: <25%. Mixed results: No majority direction is found across all outcomes.

Summary of age-responsive interventions – SDG 3

	Child health	SRH and maternal health	Mental health and well-being	Access and use of health services	Nutrition and dietary diversity	Child labour	Other health outcomes
#studies	66	17	11	5	37	6	17
% of outcomes measured are effective	42%	55%	54%	64%	48%	20%	42%
	Positive & weak, design-dependent	Positive & strong, esp. for vouchers	Positive & strong	Positive & strong	Positive & weak	Mixed results	Positive & weak

Source: Synthesis team elaboration

Notes: Positive & strong: >50% outcomes are positive and significant; Positive & weak: <50% outcomes but >than 30%; Not effective: <25%. Mixed results: No majority direction is found across all outcomes.

Table 1.

SDG 1 - Gender-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Ahmed (2024a)	MI, rural	E	●	Caregivers Infants and children (0-5)	Female caregivers	(▲)*$_{1/3}$
Ahmed (2024b)	MI, rural	E	●	Female caregivers, infants (0-5)	Female caregivers, Infants and children (0-5)	(▲)*$_{1/8}$
Chinen (2017)	Mix, Mix	SR	●	Young adults (18-34), esp. women	Young adults (18-34), esp. women	(▲)*$_{2/2}$
Leventhal (2016)	MI, rural	E	●	Adolescents (14-17)	Adolescent girls (14-17)	(▲)*$_{3/3}$
Wang (2024)	UI, rural	QE	●	Mothers or female caregivers	Mothers or female caregivers	(▲)*$_{1/1}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 2.

SDG 1 - Age-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Ahmed (2024b)	MI, rural	E	●	Female caregivers, Infants and children (0-5)	Female caregivers, Infants and children (0-5)	(▲)*$_{1/8}$
Baird (2016)	MI, N/S	E	●	Children (6-13) in 1998-2003	Young adults (19–26)	(▲)*$_{2/5}$
Gross (2020)	UI, N/S	SR	●	Young adults (18-34), adults (35-64) with disabilities	Young adults (18-34), adults (35-64) with disabilities	(▲)*$_{1/1}$
Little (2021)	Mix, N/S	SR	●	Infants and children (0-5)	HHs with Infants and children (0-5)	(▲)*$_{4/4}$
Prencipe (2022)	MI, N/S	E	●	Adolescents (14-17)	Adolescents (14-17)	(▲)*$_{2/3}$
Rosas (2017)	LI, urban	E	●	Young adults (18-34)	Young adults (18-34)	(▲)*$_{2/5}$
Tranchant (2018)	LI, rural	QE	●	Caregivers, Infants and children (0-13)	HHs with Infants and children (0-5)	(▲)*$_{1/1}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 3.

SDG 1 – Disability-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Gross (2020)	UI, N/S	SR	●	Young adults (18-34) Adults (35-64) with disabilities	Young adults (18-34) Adults (35-64) with disabilities	(▲)*$_{1/1}$
Hunt (2022)	MI, mix	SR	●	Person with disabilities	Person with disabilities	(▲)*$_{6/6}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 4.

SDG 1 – Multiple vulnerability-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Jeong (2022)	LI, mix	SR	●	HHs receiving humanitarian assistance	HHs receiving humanitarian assistance	(▲)*$_{1/2}$
Quattrochi (2020)	LI, rural	E	●	Migrants and IDPs	Migrants and IDPs	(▲)*$_{3/5}$
Tagliati (2022)	UI, rural	E	●	Adolescents (14-17) from ultra-poor households	Adolescents (14-17) from ultra-poor households	(▲)*$_{1/8}$
Tranchant (2018)	LI, rural	QE	●	Caregivers, Infants and children (0-13)	HHs with Infants (0-5)	(▲)*$_{1/1}$
Wossen (2017)	MI, rural	E	●	Farmers (18-64)	Farmers (18-64)	(▲)*$_{3/3}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 5.

SDG 2 - Gender-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Ahmed (2024a)	MI, rural	E	●	Caregivers, Infants and children (0-5)	Female caregivers	(▲)*$_{2/4}$
Ahmed (2024b)	MI, rural	E	●	Female caregivers, Infants and children (0-5)	Household	(▲)*$_{9/10}$
Doocy (2019)	LI, N/S	QE	●	Female caregivers, Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{3/3}$
Fitzsimons (2016)	LI, N/S	E	●	Female caregivers, Infants and children (0-5)	Household	(▲)*$_{3/5}$
Gelli (2018)	LI, rural	QE	●	Female caregivers Children (6-13) Adolescents (14-17)	Household	(▲)*$_{1/2}$
Visser (2020)	Mix, Mix	SR	●	Mothers or female caregivers Infants and children (0-13) Adolescents (14-17) Elderly (65+)	Households in disadvantaged and vulnerable communities	(◀▶)$_{2/2}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 6.

SDG 2 - Age-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Ahmed (2024b)	MI, rural	E	●	Female caregivers and infants (0-5)	Infants and children (0-5)	(▲)*$_{5/7}$
Doocy (2019)	LI, N/S	QE	●	Female caregivers, children under 5 years	Infants and children (0-5)	(▲)*$_{3/3}$
Fitzsimons (2016)	LI, N/S	E	●	Female caregivers and infants (0-5)	Infants and children (0-5)	(▲)*$_{4/11}$
Gelli (2018)	LI, rural	QE	●	Female caregivers, Children (6-13), Adolescents (14-17)	Household	(▲)*$_{1/2}$
Gilligan (2016)	LI, rural	E	●	Infants and children (0-5)	Infants and children (0-5)	(▲)$_{9/11}$
Jamaluddine (2022)	MI, rural	QE	●	Children (6-13)	Children (6-13)	(▲)*$_{9/16}$
Just (2022)	MI, rural	E	●	Children (6-13)	Children (6-13)	(▲)*$_{3/3}$
Kalid (2022)	LI, rural	QE	●	Infants and children (0-5)	Infants and children (0-5)	(◀▶)$_{3/3}$
Larson (2023)	MI, N/S	QE	●	Infants and children (0-5)	Infants and children (0-5)	(▲)$_{4/5}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Schreinemachers (2020)	MI, Mix	E	●	Children (6-13)	School children (aged 8–12)	(▲)*$_{4/10}$
Tranchant (2018)	LI, rural	QE	●	Caregivers and infants (0-5) Children (6-13)	Infants and children (0-5)	(▲)*$_{2/7}$
Visser (2020)	Mix, Mix	SR	●	Mothers or female caregivers Infants (0-5), Children (6-13), Adolescents (14-17), Elderly (65+)	Households in disadvantaged and vulnerable communities	(◀▶)$_{2/2}$
Wellington (2021)	UI, rural	QE	●	Children (6-13); Adults (35-64); Elderly (65+)	Household	(▲)*$_{9/9}$
Xu (2020)	UI, urban	E	●	Children (6-13)	Children (6-13)	(▲)*$_{6/14}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 7.

SDG 2 – Disability-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Wellington (2021)	UI, rural	QE	●	Children (6-13) Adults (35-64) Elderly (65+) HHs with Persons with Disabilities	HHs (HH heads)	(▲)*$_{9/9}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 8.

SDG 2– Multiple vulnerability-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Jeong (2022)	LI, mix	SR	●	HHs receiving humanitarian assistance	HHs receiving humanitarian assistance	(▲)*$_{2/2}$
Quattrochi (2020)	LI, rural	E	●	Migrants and IDPs, as well as non-displaced vulnerable host communities	Migrants and IDPs, as well as non-displaced vulnerable host communities	(▲)*$_{1/1}$
Tagliati (2022)	UI, rural	E	●	Adolescents (14-17) from ultra-poor households	Adolescents (14-17) from ultra-poor households	(▲)$_{1/1}$
Tranchant (2018)	LI, rural	QE	●	Caregivers, Infants (0-5) and children (6-13)	HHs with infants and children in conflict affected and areas	(▲)*$_{4/7}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Wellington (2021)	UI, rural	QE	●	Children (6-13) Adults (35-64) Elderly (65+)	HHs (HH heads)	(▲)*$_{9/9}$
Wossen (2017)	MI, rural	E	●	Farmers (18-64)	Farmers (18-64)	(▲)*$_{3/3}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 9.

SDG 3 - Gender-responsive: child health

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Ahmed (2023)	LI, N/S	SR	🟢	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{1/4}$
Ahmed (2024b)	MI, rural	E	🟢	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{3/12}$
Ali (2020)	MI, mix	QE	🔴	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{3/8}$
Andersson (2024)	LI, rural	QE	🟡	Mothers or female caregivers (pregnant women)	Infants and children (0-5) - Newborns	(▲)*$_{1/1}$
Das (2018)	Mix, mix	SR	🟢	Mothers or female caregivers (pregnant women)	Infants and children (0-5) Newborns	(▲)*$_{6/8}$
Doocy (2019)	LI, N/S	QE	🟡	Female caregivers, Infants and children (0-5)	Infants and children (0-5)	(▲)$_{3/4}$
Escher (2024)	Mix, mix	SR	🟢	All groups	Infants and children (0-13)	(◀▶)$_{1}$
Fitzsimon (2016)	LI, N/S	E	🟢	Female caregivers, Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{6/8}$
Gelli (2018)	LI, rural	QE	🟡	Female caregivers Children (6-13) Adolescents (14-17)	Children (6-13)	(◀▶)$_{1}$
Lassi (2021)	Mix, mix	SR	🟢	Infants and children (0-5) Adolescents (14-17) Adults (35-64)	Infants and children (0-5)	(▲)$_{2}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Mbuagbaw (2015)	Mix, mix	SR	●	Mothers or female caregivers Infants and children (0-5) Young adults (18-34) Adults (35-64)	Infants and children (0-5)	(▲)*$_{2/2}$
Narayanan (2019)	MI, mix	QE	●	Children (6-13) Women	Children (6-13)	(▲)*$_{1/1}$
Neelsen (2021)	MI, mix	SR	●	Mothers or female caregivers	Infants and children (0-5)	(▲)*$_{1/1}$
Ojha (2020)	Mix, mix	SR	●	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{1/6}$
Polec (2015)	Mix, rural	SR	●	Household members	Infants and children (0-5)	(▲)*$_{1/1}$
Salam (2020)	Mix, rural	SR	●	Children (6-13) Adolescents (14-17); Especially girls for both groups	Adolescents (14-17), especially girls	(▲)$_{4/5}$
Schlossman (2017)	MI, rural	E	●	Mothers or female caregivers, Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{4/5}$
Visser (2020)	Mix, mix	SR	●	Mothers or female caregivers Infants and children (0-13) Adolescents (14-17) Elderly (65+)	Infants and children (0-5)	(▲)$_{2}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 10.

SDG 3 - Gender-responsive: sexual, reproductive, and maternal health

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Alfonso (2015)	LI, rural	QE	🟢	Mothers or female caregivers (pregnant women)	Mothers or female caregivers (pregnant women)	(▲)*$_{3/3}$
Ali (2019)	MI, mix	QE	🟡	Female: Young adults (18-34) and adults (35-64)	Female: Young adults (18-34) and adults (35-64)	(▲)*$_{3/3}$
Ali (2020)	MI, mix	QE	🔴	Female: Young adults (18-34) and adults (35-64)	Female: Young adults (18-34) and adults (35-64)	(▲)*$_{2/3}$
Anukriti (2022)	MI, rural	E	🟢	Mothers or female caregivers Young adults (18-34)	Mothers or female caregivers. Young adults (18-34)	(▲)*$_{3/5}$
Austrian (2020)	MI, mix	E	🟢	Adolescent girls (13-17)	Adolescent girls (13-17)	(▲)*$_{2/8}$
Austrian (2021)	MI, mix	E	🟡	Girls (6-13 at the beginning of the programme)	Adolescent girls (13-17 at the end of programme)	(▲)*$_{1/2}$
Azmat (2016)	MI, rural	QE	🟡	Mothers or female caregivers	Mothers or female caregivers	(▲)*$_{3/3}$
Bajracharya (2016)	MI, N/S	QE	🟡	Female young adults (18-34)	Female young adults (18-34)	(▲)*$_{2/8}$
Benjamin (2017)	MI, mix	SR	🟡	Mothers or female caregivers	Mothers or female caregivers	(▲)*$_{4/4}$
Chanda (2017)	MI, N/S	E	🟢	Female sex workers: Adults (35-64)	Female sex workers: Adults (35-64)	(▲)*$_{1/11}$
Das (2018)	Mix, mix	SR	🟢	Mothers or female caregivers (pregnant women)	Mothers or female caregivers (pregnant women)	(◀▶)$_{2}$
Gavine (2022)	Mix, mix	SR	🟢	Mothers or female caregivers	Mothers or female caregivers	(▲)$_{1}$
Lassi (2021)	Mix, mix	SR	🟢	Infants and children (0-5) - Newborns Adolescents (14-17) Adults (35-64)	Adolescents (14-17) Adults (35-64)	(▲)*$_{4/6}$
Massavon (2017)	Li, rural	QE	🟡	Mothers or female caregivers	Mothers or female caregivers (pregnant women)	(▲)*$_{3/3}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Mbuagbaw (2015)	Mix, mix	SR	●	Mothers or female caregivers Infants and children (0-5) Young adults (18-34) Adults (35-64)	Mothers or female caregivers (pregnant women)	(▲)*$_{3/7}$
Mehboob (2015)	MI, N/S	SR	●	Mothers or female caregivers Young adults (18-34)	Mothers or female caregivers (pregnant women)	(▲)*$_{4/4}$
Neelsen (2021)	MI, mix	SR	●	Mothers or female caregivers	Mothers or female caregivers	(▲)*$_{2/5}$
Obare (2016)	LI, mix	QE	●	Mothers or female caregivers	Mothers or female caregivers	(◀▶)$_{7}$
Schlossman (2017)	MI, rural	E		Mothers or female caregivers Infants and children (0-5)	Mothers or female caregivers (pregnant women)	(▲)*$_{1/3}$
Stark (2022)	LI, rural	SR	●	Adolescents (14-17), esp. female	Adolescents (14-17), esp. female	(▲)*$_{1/1}$
Till (2015)	MI, rural	SR	●	Mothers or female caregivers Children (6-13)	Mothers or female caregivers (pregnant women)	(◀▶)$_{2}$
Watt (2015)	MI, mix	E	●	Mothers or female caregivers Infants and children (0-5)	Mothers or female caregivers (pregnant women)	(▲)*$_{3/6}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 11.

SDG 3 - Gender responsive: mental health and well-being

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Banerjee (2018)	MI, rural	E and QE	●	Household members	Household members (effects reported by gender)	(◂▸)$_{1}$
Leventhal (2016)	MI, rural	E	●	Adolescents (14-17)	Adolescent girls (14-17)	(▲)*$_{1/2}$
Stark (2022)	LI, rural	SR	●	Adolescents (14-17)	Adolescent girls (14-17)	(▲)*$_{1/1}$
Whillans (2022)	MI, urban	E	●	Mothers or female caregivers	Mothers or female caregivers (working mothers)	(▲)$_{3}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◂▸ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 12.

SDG 3 - Gender responsive: access and use of health services

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Polec (2015)	Mix, rural	SR	●	Household members	Household members	(▲)*$_{2/2}$
Watt (2015)	MI, mix	E	●	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)$_{3}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 13.

SDG 3 - Gender-responsive: nutrition and dietary diversity

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Adelman (2019)	LI, rural	E	●	Infants and children (0-5) Children (6-13) Women (18-35)	Infants and children (0-5) Children (6-13) Women (18-35)	(▲)*$_{4/4}$
Das (2018)	Mix, mix	SR	●	Mothers or female caregivers (pregnant women)	Mothers or female caregivers (pregnant women)	(▼)*$_{1/1}$
Doocy (2019)	LI, N/S	QE	●	Female caregivers, Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{2/2}$
Fitzsimons (2016)	LI, N/S	E	●	Female caregivers and Infants and children (0-5)	Female caregivers	(▲)*$_{3/8}$
Gelli (2018)	LI, rural	QE	●	Female caregivers Children (6-13) Adolescents (14-17)	Children (6-13)	(▲)*$_{1/5}$
Hoddinott (2018)	MI, rural	E	●	Mothers or female caregivers Infants and children (0-5)	Mothers or female caregivers	(▲)*$_{2/2}$
Kim (2023b)	LI, N/S	E	●	Children (6-13), girls	Children (6-13), girls only	(▲)*$_{6/7}$
Lassi (2021)	Mix, mix	SR	●	Infants and children (0-5) Newborns Adolescents (14-17) Adults (35-64)	Adolescents (14-17)	(◀▶)$_{1}$
Leventhal (2016)	MI, rural	E	●	Adolescents (14-17)	Adolescent girls (14-17)	(▲)$_{1}$
Salam (2020)	Mix, rural	SR	●	Children (6-13) Adolescents (14-17); Esp. girls for both groups	Adolescents (14-17)	(▲)$_{8}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 14.

SDG 3 - Gender-responsive: child labour

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Aurino (2019)	LI, rural	QE	•	Children (6-13), esp. girls	Children (6-13)	$(◀▶)_2$
Gelli (2018)	LI, rural	QE	•	Female caregivers Children (6-13) Adolescents (14-17)	Children (6-13)	$(◀▶)_3$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: • low quality; • medium quality; • high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 15.

SDG 3 - Gender-responsive: other health outcomes

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Banerjee (2018)	MI, rural	E and QE	●	Households	Household members (effects reported by gender)	(◄►)$_{4}$
Leventhal (2016)	MI, rural	E	●	Adolescents (14-17)	Adolescent girls (14-17)	(▲)*$_{4/4}$
Narayanan (2019)	MI, mix	QE	●	Children (6-13) Women	Women	(◄►)$_{2}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◄► (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 16.

SDG 3 - Age-responsive - child health

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Aboud (2017)	LI, Rural	E	●	Infants and children (0-5)	Children aged 4 to 6 years	(▲)*$_{5/5}$
Adom (2020)	Mix, Mix	SR	●	Children (6-13) Adolescents (14-17)	Elementary school children and Adolescents in grades 4 to 6	(▲)*$_{7/9}$
Ahmed (2021)	MI, Rural	E	●	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{4/11}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Ahmed (2023)	LI, N/S	SR	●	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{1/4}$
Ahmed (2024b)	MI, Rural	E	●	Female caregivers and Infants and children (0-5)	Infants and children (0-5)	(◀▶)$_{3/12}$
Ali (2020)	MI, Mix	QE	●	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{3/8}$
Ali (2024)	MI, Rural	E	●	Infants and children (0-5)	Infants and children (0-5)	(◀▶)*$_{4/9}$
Andrade (2018)	UI, Mix	SR	●	Children (6-13)	Children (6-13)	(▲)*$_{2/2}$
Andrew (2020)	MI, Urban	E	●	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{5/6}$
Angawi (2021)	UI, N/S	SR	●	Children (6-13) Adolescents (14-17)	Children and adolescents (4-18)	(◀▶)*$_{3/3}$
Baird (2016)	MI, N/S	E	●	Children (6-13)	Children (6-13)	(▲)*$_{2/4}$
Berry (2020)	MI, Rural	QE	●	Children (6-13)	Children (6-13)	(▲)$_{2}$
Bidira (2022)	LI, Mix	QE	●	Infants and children (0-5)	Preschool children	(▲)*$_{2/3}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Carducci (2020)	UI, Urban	SR	🔴	Children (6-13) Adolescents (14-17)	School-aged children	(▲)*$_{1}$
Chen (2020)	UI, Rural	E	🟢	Children (6-13)	School children (grades 4 and 5)	(▲)*$_{1/3}$
Clasen (2015)	Mix, Mix	SR	🟢	All age groups	All age groups	(▲)*$_{1/2}$
Cliffer (2023)	LI, Mix	E	🟡	Children (6-13) Adolescents (14-17)	Children and adolescents 10-18	(▲)$_{2}$
Das (2019)	Mix, Rural	SR	🟢	Infants and children (0-5)	Infants and children 6-23 months	(▲)*$_{10/13}$
Doocy (2019)	LI, N/S	QE	🟡	Female caregivers, Infants and children (0-5)	Infants and children (0-5)	(▲)$_{3/4}$
Dulal (2018)	MI, Rural	E	🟢	Infants and children (0-5)	Infants and children (0-5)	(▼)$_{9}$
Eaton (2019)	Mix, Rural	SR	🟡	Infants and children (0-5)	Infants and children (0-5)	(▼)$_{3}$
Escher (2024)	Mix, Mix	SR	🟢	All groups	Infants and children (0-5) Children (6-13)	(◀▶)$_{1}$
Fang (2022)	UI, Rural	QE	🟢	Children (6-13)	Children (6-13)	(▲)*$_{3/5}$
Fitzsimons (2016)	LI, N/S	E	🟢	Female caregivers and Infants and children (0-5)	Infants and children and children (0-5)	(▲)*$_{4/5}$
Gelli (2018)	LI, Rural	QE	🟡	Female caregivers Children (6-13) Adolescents (14-17)	Children (6-13)	(◀▶)$_{1}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Gelli (2019)	MI, Mix	E	●	Children (6-13)	School-aged Children (5-15)	(▲)*$_{2/2}$
Gilligan (2016)	LI, Rural	E	●	Infants and children (0-5)	Infants and children (0-5)	(▼)*$_{2/10}$
Imdad (2022)	Mix, Mix	SR	●	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{3/8}$
Kalid (2022)	LI, Rural	QE	●	Infants and children (0-5)	Infants and children (0-5)	(▼)$_{5/6}$
Kim(2023) [a]	MI, N/S	E	●	Children (6-13)	Children (6-13)	(▲)*$_{1/3}$
Klingberg (2019)	Mix, Mix	SR	●	Infants and children (0-5) Children (6-13) Adolescents (14-17)	Children (6-13) Adolescents (14-17)	(▲)*$_{2/2}$
Kristjansson (2015)	MI, Mix	SR	●	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{5/8}$
Lassi (2021)	Mix, Mix	SR	●	Adolescents (14-17) Adults (35-64)	Infants and children (0-5) - Newborns	(▲)$_{2}$
Little (2021)	Mix, N/S	SR	●	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{4/10}$
Ma (2015)	UI, Rural	E	●	Children (6-13)	Children (6-13)	(▲)*$_{1/1}$
Mazumder (2019)	MI, Rural	E	●	Infants and children (0-5)	Newborns	(▲)*$_{6/9}$
Mbuagbaw (2015)	Mix, Mix	SR	●	Mothers or female caregivers Infants and children (0-5) Young adults (18-34) Adults (35-64)	Infants and children (0-5) - Newborns	(▲)*$_{2/2}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Mohapatra (2023)	MI, Mix	QE	●	Children (6-13)	Children (6-13)	(▲)*$_{2/2}$
Narayanan (2019)	UI, Mix	QE	●	Children (6-13) Women	Children (6-13)	(▲)*$_{1/1}$
Neelsen (2021)	MI, Mix	SR	●	Mothers or female caregivers	Infants and children (0-5)	(▲)*$_{1/1}$
Ochoa (2017)	UI, Urban	E	●	Children (6-13) Adolescents (14-17)	School children aged 12-14	(▲)*$_{1/2}$
Ojha (2020)	Mix, Mix	SR	●	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{1/6}$
Patel (2018)	MI, Mix	QE	●	Children (6-13) Adolescents (14-17)	children and Adolescents (11-18)	(◀▶)$_{2}$
Polec (2015)	Mix, Rural	SR	●	Household members	Infants and children (0-5)	(▲)*$_{1/1}$
Pongutta (2022)	Mix, Urban	SR	●	Children (6-13)	Children (6-13)	(▲)*$_{1/1}$
Ren (2023)	UI, Rural	QE	●	Children (6-13) Adolescents (14-17)	School children aged 6-16	(▲)*$_{1}$
Reyes-Morales (2016)	UI, Urban	E	●	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{1}$
Salam (2020)	Mix, Rural	SR	●	Children (6-13) Adolescents (14-17); Especially girls for both groups	Adolescents (14-17), especially girls	(▲)$_{4/5}$
Saucedo-Delgado (2023)	UI, Rural	QE	●	Children (6-13)	Children (6-13)	(▼)*$_{1/2}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Saville (2018)	LI, Rural	E	🟢	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{2/3}$
Schlossman (2017)	MI, Rural	E	🔴	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{4/5}$
Schreinemachers (2017)	MI, Rural	E	🟡	Children (6-13)	School children in grades 6 and 7	(▲)$_{1}$
Singhal (2021)	UI, Mix	SR	🟢	Children (6-13)	School children aged 4 to 12 years old	(▲)*$_{1/3}$
Tam (2020)	Mix, Mix	SR	🟢	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{7/12}$
Tranchant (2018)	LI, Rural	QE	🟢	Caregivers and Infants and children (0-5) Children (6-13)	Infants and children and children (0-5)	(▲)*$_{1}$
Trenouth (2018)	LI, Rural	E	🟡	Infants and children (0-5)	Infants and children aged 4 to 48 months	(▲)*$_{1/2}$
Visser (2020)	Mix, Mix	SR	🟡	Mothers or female caregivers Infants and children (0-5) Children (6-13) Adolescents (14-17) Elderly (65+)	Infants and children, children, and adolescents (0-17) in disadvantaged and vulnerable households	(▲)$_{2}$
von Grafenstein (2023)	MI, N/S	E	🔴	Children (6-13)	Children (6-13)	(▲)*$_{2/2}$
Wang (2019)	UI, Rural	QE	🔴	Children (6-13)	Children (6-13)	(◀▶)$_{4}$
Wang (2022a)	UI, Rural	QE	🟢	Children (6-13)	Children (6-13)	(▲)*$_{1/3}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Wang (2022c)	UI, Rural	E	●	Children (6-13) Adolescents (14-17)	Children (6-13) Adolescents (14-17)	(▲)*$_{1}$
Welch (2016)	Mix, Mix	SR	●	Infants and children (0-5) Children (6-13) Adolescents (14-17)	Infants and children (0-5) Children (6-13) Adolescents (14-17)	(▲)$_{5}$
Xu (2017)	UI, Urban	E	●	Children (6-13)	Children (6-13)	(◀▶)*$_{8}$
Zhang (2024)	UI, Rural	E	●	Infants and children (0-5)	Preschool children	(▲)*$_{1/2}$
Zhao (2020)	UI, Rural	E	●	Children (6-13)	Children (6-13)	(▲)*$_{1}$
Zhao (2024)	UI, Urban	QE	●	Children (6-13) Adolescents (14-17)	Children (6-13) Adolescents (14-17)	(▲)*$_{4/4}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 17.

SDG 3 - Age responsive- sexual, reproductive, and maternal health

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Ahmed (2021)	MI, rural	E	•	Infants and children (0-5)	Mothers of young children	(◄►)$_{1}$
Ali (2019)	MI, mix	QE	•	Female: Young adults (18-34) and adults (35-64)	Female: Young adults (18-34) and adults (35-64)	(▲)*$_{3/3}$
Ali (2020)	MI, mix	QE	•	Female: Young adults (18-34) and adults (35-64)	Female: Young adults (18-34) and adults (35-64)	(▲)*$_{2/3}$
Anukriti (2022)	MI, rural	E	●	Mothers or female caregivers Young adults (18-34)	Mothers or female caregivers. Young adults (18-34)	(▲)*$_{3/5}$
Baird (2016)	MI, N/S	E	•	Children (6-13)	Pregnant women	(▲)*$_{1}$
Belaid (2016)	Mix, N/S	SR	●	Adolescents (14-17) Young adults (18-34) Adults (35-64)	Adolescents (14-17) Young adults (18-34) Adults (35-64)	(▲)*$_{1/3}$
Chong (2020)	UI, Urban	E	●	Adolescents (14-17)	Adolescents (14-17)	(▲)*$_{7/7}$
Dulal (2018)	MI, Rural	E	●	Infants and children (0-5)	Pregnant women	(▲)$_{2}$
Gavine (2022)	Mix, mix	SR	●	Mothers or female caregivers	Mothers or female caregivers	(▲)$_{1}$
Lassi (2021)	Mix, mix	SR	●	Infants and children (0-5) - Newborns Adolescents (14-17) Adults (35-64)	Adolescents (14-17) Adults (35-64)	(▲)*$_{4/6}$
Mbuagbaw (2015)	Mix, mix	SR	●	Mothers or female caregivers Infants and children (0-5) Young adults (18-34) Adults (35-64)	Mothers or female caregivers (pregnant women)	(▲)*$_{3/7}$
Neelsen (2021)	MI, mix	SR	●	Mothers or female caregivers	Mothers or female caregivers	(▲)*$_{2/5}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Patel (2018)	MI, Mix	QE	•	Children (6-13) Adolescents (14-17)	Children and Adolescents (11-18)	(◄►)$_{3}$
Rose-Clarke (2019)	Mix, N/S	SR	•	Adolescents (14-17)	Persons aged 10 to 19	(◄►)*$_{5}$
Schlossman (2017)	MI, rural	E	•	Mothers or female caregivers Infants and children (0-5)	Mothers or female caregivers (pregnant women)	(▲)*$_{1/3}$
Till (2015)	MI, rural	SR	●	Mothers or female caregivers Children (6-13)	Mothers or female caregivers (pregnant women)	(◄►)$_{2}$
Watt (2015)	MI, mix	E	•	Mothers or female caregivers Infants and children (0-5)	Mothers or female caregivers (pregnant women)	(▲)*$_{3/6}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◄► (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 18.

SDG 3 - Age-responsive - mental health and well-being

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Banerjee (2018)	MI, rural	E and QE	●	Household members	Household members (effects reported by gender)	(◀▶)$_{1}$
Bangpan (2024)	Mix, N/S	SR	●	Children (6-13)	Children (6-13)	(▲)$_{4}$
Domínguez-Viera (2024)	UI, Urban	E	●	Children (6-13)	School children 9-12	(▲)*$_{3/6}$
Fang (2022)	UI, Rural	QE	●	Children (6-13)	Children (6-13)	(▲)*$_{2/5}$
Jamaluddine (2022)	MI, rural	QE	●	Children (6-13)	Children (6-13)	(▲)*$_{9/16}$
Li (2024)	UI, Rural	E	●	Infants and children (0-5)	Infants and children (0-5)	(▲)$_{2}$
Liu (2019)	UI, Rural	E	●	Children (6-13)	Children (6-13)	(▲)*$_{8/9}$
Prencipe (2022)	LI, N/S	E	●	Adolescents (14-17)	Youth (14-19)	(▲)*$_{2/4}$
Purgato (2018)	LI, Urban	SR	●	Adolescents (14-17)	Adolescents (14-18) exposed to traumatic events	(▲)*$_{3/3}$
Quattrochi (2020)	LI, rural	E	●	Migrants and IDPs as well as non-displaced vulnerable host communities	Adults (35-64)	(▲)*$_{2/3}$
Rose-Clarke (2019)	Mix, N/S	SR	●	Adolescents (14-17)	Persons aged 10 to 19	(▲)*$_{2/4}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 19.

SDG 3 - Age-responsive - access and use of health services

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Makata (2021)	MI, Mix	E	●	Children (6-13)	Children aged 8 to 12 years	(▲)*$_{2/3}$
Polec (2015)	Mix, rural	SR	●	Household members	Household members	(▲)*$_{2/2}$
Wang (2022b)	UI, Rural	E	●	Children (6-13)	Children (6-13)	(▲)*$_{2/2}$
Watt (2015)	MI, mix	E	•	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)$_{3}$
Zhang (2022)	UI, rural	E	●	Children (6-13)	Children (6-13)	(▲)*$_{1}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 20.

SDG 3 - Age-responsive - nutrition and dietary diversity

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Aboud (2017)	LI, Rural	E	•	Infants and children (0-5)	Children aged 4 to 6 years	(▲)*$_{2/2}$
Adelman (2019)	LI, Rural	E	•	Infants and children (0-5) Children (6-13) Women (18-35)	Infants and children (0-5) Children (6-13) Women (18-35)	(▲)*$_{4/4}$
Adom (2020)	Mix, Mix	SR	•	Children (6-13) Adolescents (14-17)	Elementary school children and Adolescents in grades 4 to 6	(▲)*$_{4/4}$
Andrade (2018)	UI, Mix	SR	•	Children (6-13)	Children (6-13)	(▲)*$_{2/2}$
Baird (2016)	MI, N/S	E	•	Children (6-13)	Children (6-13)	(▲)*$_{1/1}$
Baliki (2023)	MI, Rural	E	•	Children (6-13)	Children (6-13)	(▲)*$_{1/5}$
Chen (2020)	UI, Rural	E	•	Children (6-13)	School children (grades 4 and 5)	(▼)*$_{1/3}$
Domínguez-Viera (2024)	UI, Urban	E	•	Children (6-13)	School children 9-12	(▲)$_{2/3}$
Doocy (2019)	LI, N/S	QE	•	Female caregivers, Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{2/2}$
Fang (2022)	UI, Rural	QE	•	Children (6-13)	Children (6-13)	(▲)*$_{2/4}$
Fitzsimons (2016)	LI, N/S	E	●	Female caregivers and Infants and children (0-5)	caregivers	(▲)*$_{3/8}$
Gelli (2018)	LI, rural	QE	•	Female caregivers Children (6-13) Adolescents (14-17)	Children (6-13)	(▲)*$_{1/5}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Gelli (2020)	LI, Rural	E	•	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{5/11}$
Hoddinott (2018)	MI, rural	E	•	Mothers or female caregivers Infants and children (0-5)	Mothers or female caregivers	(▲)*$_{2/2}$
Imdad (2022)	Mix, Mix	SR	•	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{1/1}$
Kalid (2022)	LI, rural	QE	•	Infants and children (0-5)	Household Infants and children (0-5)	(▼)$_{1}$
Kim (2023)	LI, N/S	E	•	Children (6-13), girls	Children (6-13), girls only	(▲)*$_{6/7}$
Klingberg (2019)	Mix, Mix	SR	•	Infants and children (0-5) Children (6-13) Adolescents (14-17)	Children (6-13) Adolescents (14-17)	(▲)$_{1}$
Larson (2023)	MI, N/S	QE	●	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{2/5}$
Lassi (2021)	Mix, mix	SR	•	Infants and children (0-5) Adolescents (14-17) Adults (35-64)	Adolescents (14-17)	(◀▶)$_{1}$
Lin (2016)	UI, N/S	E	•	Infants and children (0-5)	Preschoolers (3-6 years old)	(▲)*$_{1/1}$
Mahajan (2022)	MI, Urban	E	•	Adolescents (14-17)	Adolescents (14-17)	(▲)*$_{2/3}$
Nguyen (2021)	MI, Urban	E	•	Children (6-13) and caregivers	Children (8-10)	(▲)*$_{2/3}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Ochoa (2017)	UI, Urban	E	•	Children (6-13) Adolescents (14-17)	School children aged 12-14	(◂▸)*$_{5}$
Reyes-Morales (2016)	UI, Urban	E	•	Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{2/2}$
Salam (2020)	Mix, Rural	SR	•	Children (6-13) Adolescents (14-17); Especially girls for both groups	Adolescents (14-17)	(▲)$_{8}$
Saraf (2015)	MI, Rural	E	•	Children (6-13) and caregivers	School children in grades 6 and 7	(▲)*$_{1/2}$
Saucedo-Delgado (2023)	UI, Rural	QE	•	Children (6-13)	Children (6-13)	(◂▸)*$_{5/7}$
Schreinemachers (2017)	MI, Rural	E	•	Children (6-13)	School children in grades 6 and 7	(▲)*$_{2/3}$
Schreinemachers (2019)	LI, N/S	E	•	Children (6-13) Adolescents (14-17)	School children 8 to 14 years	(▲
Schreinemachers (2020)	MI, Mix	E	•	Children (6-13)	Schoolchildren (aged 8–12)	(▲)*$_{3/3}$
Teng (2018)	UI, Mix	E	•	Children (6-13) Adolescents (14-17)	Children (6-13) Adolescents (14-17)	(▲)*$_{4/6}$
Tranchant (2018)	LI, Rural	QE	•	Caregivers, Infants and children (0-13)	Infants and children (0-5)	(▲)$_{1}$
Wang (2022c)	UI, Rural	E	•	Children (6-13) Adolescents (14-17)	Parents and households	(▲)*$_{1/2}$
Xu (2020)	UI, Urban	E	•	Children (6-13)	children aged 7-13 years	(▲)*$_{6/14}$
Zhang (2024)	UI, Rural	E	•	Infants and children (0-5)	Preschool children	(▲)*$_{2/2}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Zhao (2024)	UI, Urban	QE	•	Children (6-13) Adolescents (14-17)	Children (6-13) Adolescents (14-17)	(▲)*$_1$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 21.

SDG 3 - Age-responsive - child labour

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Aurino (2019)	LI, rural	QE	•	Children (6-13), esp. girls	Children (6-13)	(◀▶)$_2$
Baryshnikova (2019)	UI, Mix	QE	•	Children (6-13)	Children (6-13)	(▲)$_1$
Dammert (2018)	Mix, Mix	SR	•	Children (6-13)	Children (6-13)	(▲)*$_1$
Gelli (2018)	LI, rural	QE	•	Female caregivers Children (6-13) Adolescents (14-17)	Children and adolescents (6-17) in conflict-affected areas	(◀▶)*$_1$
Morgan (2015)	UI, N/S	SR	•	Children (6-13)	Children (6-13)	(▲)*$_1$
Tagliati (2022)	UI, rural	E	•	Adolescents (14-17) from ultra-poor households	Adolescents (14-17) from ultra-poor households	(◀▶)$_4$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 22.

SDG 3 - Age-responsive - other health outcomes

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Adom (2020)	Mix, Mix	SR	●	Children (6-13) Adolescents (14-17)	Elementary school children and Adolescents in grades 4 to 6	(▲)*$_{2/3}$
Andrew (2020)	MI, Urban	E	●	Mothers or female caregivers Infants and children (0-5)	Infants and children (0-5)	(▲)*$_{1/1}$
Baliki (2023)	MI, Rural	E	●	Children (6-13)	Children (6-13)	(▲)*$_{1/1}$
Banerjee (2018)	MI, rural	E and QE	●	Households	Household members (effects reported by gender)	(◀▶)$_{4}$
Clasen (2015)	Mix, Mix	SR	●	All age groups	All age groups	(▲)*$_{1/1}$
Ellermeijer (2023)	LI, N/S	SR	●	Infants and children (0-5) Children (6-13)	Children (6-13)	(▲)*$_{2/2}$
Little (2021)	Mix, N/S	SR	●	Infants and children (0-5)	Infants and children (0-5)	(▲)$_{2}$
Mahajan (2022)	MI, Urban	E	●	Adolescents (14-17)	Adolescents (14-17)	(▲)$_{4}$
Makata (2021)	MI, Mix	E	●	Children (6-13)	Children aged 8 to 12 years	(▼)$_{1}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Narayanan (2019)	MI, mix	QE	●	Children (6-13) Women	Women	(◄►)$_{2}$
Ning (2019)	UI, N/S	E	●	Infants and children (0-5) and caregivers	Preschoolers aged 3-6 years and caregivers	(◄►)$_{3}$
Patel (2018)	MI, Mix	QE	●	Children (6-13) Adolescents (14-17)	Children and Adolescents (11-18)	(▼)$_{1}$
Peralta (2022)	UI, mix	SR	●	Adolescents (14-17)	Adolescents (14-17)	(▼)*$_{3/3}$
Rose-Clarke (2019)	Mix, N/S	SR	●	Adolescents (14-17)	Persons aged 10 to 19	(▲)*$_{3/3}$
Saraf (2015)	MI, Rural	E	●	Children (6-13) and caregivers	Caregivers	(▲)$_{1}$
Wang (2022a)	UI, Rural	QE	●	Children (6-13)	Children (6-13)	(▲)*$_{1/3}$
Wellington (2021)	UI, rural	QE	●	Children (6-13) Adults (35-64) Elderly (65+)	Households (HH heads responded for the HHs)	(▲)*$_{1/1}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◄► (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 23.

SDG 3 - Disability-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Wang (2022b)	UI, rural	E	●	Children (6-13) with disabilities (myopia)	Children (6-13) with disabilities (myopia)	(▲)*$_{2/2}$
Wellington (2021)	UI, rural	QE	●	Children (6-13) Adults (35-64) Elderly (65+)	Households (HH heads responded for tHHs)	(▲)$_{1/1}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 24.

SDG 3– Multiple vulnerability-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Ahmed (2021)	MI, rural	E	●	Mothers or female caregivers Infants and children (0-5)	Mothers or female caregivers Infants and children (0-5)	(◀▶)$_{6}$
Jayawardana (2021)	UI, mix	QE	●	Children (6-13) Adults (35-64)	Children (6-13) Adults (35-64)	(◀▶)$_{2}$
Quattrochi (2020)	LI, rural	E	●	Migrants and IDPs, as well as non-displaced vulnerable host communities	Children (6-13) Adults (35-64)	(▲)*$_{3/10}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Tagliati (2022)	UI, rural	E	●	Adolescents (14-17) from ultra-poor households	Adolescents (14-17) from ultra-poor households	(◀▶) $_{4/4}$
Tranchant (2018)	LI, rural	QE	●	Caregivers, Infants and children (0-5) Children (6-13)	Infants and children (0-5) Children (6-13)	(▲)* $_{2/2}$
Watt (2015)	MI, mix	E	●	Mothers or female caregivers Infants and children (0-5)	Mothers or female caregivers Infants and children (0-5)	(▲)* $_{3/9}$
Wellington (2021)	UI, rural	QE	●	Children (6-13) Adults (35-64) Elderly (65+)	Households (HH heads responded for HHs)	(▲) $_{1/1}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 25.

SDG 4 - Gender-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Aurino (2019)	LI, rural	QE	●	Children (6-13), esp. girls	Children (6-13)	(◀▶) $_{3}$
Azomahou (2019)	MI, rural	E, QE	●	Children (6-13), esp. girls	Children (6-13)	(▲)* $_{4/7}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Cheung (2015)	MI, rural	E	●	Children (6-13), esp. girls	Children (6-13)	(▲)*$_{2/2}$
Gelli (2018)	LI, rural	QE	●	Children (6-13) and Adolescents (14-17), esp. female	Children (6-13) and Adolescents (14-17)	(◀▶)$_{3}$
Kim (2023b)	LI, N/S	E	●	Children (6-13), girls only	Children (6-13), girls only	(▲)*$_{1/4}$
Stark (2022)	LI, rural	SR	●	Adolescents (14-17), esp. female	Adolescents (14-17), esp. female	(▲)*$_{1/1}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 26.

SDG 4 - Age-responsive: access to education

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Aurino (2019)	LI, rural	QE	●	Children (6-13), esp. girls	Children (6-13)	(◀▶)$_{2}$
Azomahou (2019)	MI, rural	E, QE	●	Children (6-13), esp. girls	Children (6-13)	(◀▶)$_{4}$
Baird (2016)	MI, N/S	E	●	Children (6-13) in 1998-2003	Young adults (19–26)	(▲)*$_{2/5}$
Baryshnikova (2019)	MI, Mix	QE	●	Children (6-13)	Children (6-13)	(▲)$_{1}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Cheung (2015)	MI, rural	E	●	Children (6-13), esp. girls	Children (6-13)	(▲)*$_{2/2}$
Dammert (2018)	Mix, mix	SR	●	Children (6-13)	Children (6-13)	(▲)*$_{1/1}$
Gelli (2018)	LI, rural	QE	●	Children (6-13) and Adolescents (14-17), esp. female	Children (6-13) and Adolescents (14-17)	(◀▶)$_{2}$
Gilligan (2016)	LI, rural	E	●	Infants (0-5)	Infants (0-5)	(▲)$_{1}$
Jamaluddine (2022)	MI, rural	QE	●	Children (6-13)	Children (6-13)	(▲)*$_{1/2}$
Kaur (2017)	MI, mix	QE	●	Children (6-13)	Children (6-13)	(▲)*$_{1/1}$
Manea (2020)	LI, rural	QE	●	Children (6-13)	Children (6-13)	(▲)*$_{4/4}$
Morgan (2015)	UI, N/S	SR	●	Children (6-13)	Adolescents (14-17)	(▲)*$_{4/5}$
Prencipe (2022)	MI, N/S	E	●	Adolescents (14-17)	Adolescents (14-17)	(▲)$_{1}$
Wall (2022)	Mix, rural	SR	●	Children (6-13)	Children (6-13)	(▲)*$_{3/4}$
Welch (2016)	Mix, mix	SR	●	Infants (0-5), Children (6-13) Adolescents (14-17)	Infants (0-5), Children (6-13) Adolescents (14-17)	(▲)*$_{1/2}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 27.

SDG 4 - Age-responsive: learning and achievement

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Aboud (2017)	LI, rural	E	🟡	Children aged 4 to 6 years	Children aged 4 to 6 years	(▲)*$_{1/3}$
Andrade (2018)	UI, mix	SR	🟡	Children (6-13)	Children (6-13)	(▲)*$_{1/1}$
Andrew (2020)	MI, urban	E	🟢	Mothers or female caregivers; Infants (0-5)	Infants (0-5)	(▲)*$_{1/4}$
Aurino (2019)	LI, rural	QE	🟡	Children (6-13), esp. girls	Children (6-13)	(◀▶)$_{1}$
Azomahou (2019)	MI, rural	E, QE	🟡	Children (6-13), esp. girls	Children (6-13)	(▲)*$_{3/3}$
Baird (2016)	MI, N/S	E	🟢	Children (6-13) in 1998-2003	Young adults (19–26)	(▲)$_{2/2}$
Chakraborty (2016)	UI, mix	QE	🟢	Children (6-13)	Children (6-13)	(▲)*$_{2/2}$
Dammert (2018)	Mix, mix	SR	🟡	Children (6-13)	Children (6-13)	(▲)*$_{1/1}$
Fang (2022)	UI, rural	QE	🟢	Children (6-13)	Children (6-13)	(▲)*$_{1/1}$
Gelli (2018)	LI, rural	QE	🟡	Children (6-13) and Adolescents (14-17), esp. female	Children (6-13) and Adolescents (14-17)	(◀▶)$_{1}$
Gilligan (2016)	LI, rural	E	🟢	Infants (0-5)	Infants (0-5)	(▼)*$_{2/5}$
Jamaluddine (2022)	MI, rural	QE	🟡	Children (6-13)	Children (6-13)	(▲)$_{1}$
Kim (2023b)	LI, N/S	E	🟡	Children (6-13), girls only	Children (6-13), girls only	(▲)*$_{1/4}$
Little (2021)	Mix, N/S	SR	🟢	Infants (0-5)	Infants (0-5)	(▲)$_{1}$
Mani (2020)	MI, rural	QE	🟢	Children (6-13) Adults (18-35)	Children (6-13)	(▲)*$_{5/5}$
Metwally (2020)	MI, rural	QE	🔴	Children (6-13)	Children (6-13)	(▲)*$_{5/6}$
Morgan (2015)	UI, N/S	SR	🔴	Children (6-13)	Children (6-13)	(▲)*$_{1/1}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Nguyen (2021)	MI, rural	E	🔴	Mothers or female caregivers Children (6-13)	Children (6-13)	(▲)$_{1}$
Rosas (2017)	LI, urban	E	🟡	Young adults (18-34)	Young adults (18-34)	(▲)*$_{2/2}$
Rose-Clarke (2019)	Mix, N/S	SR	🟡	Adolescents (14-17)	Adolescents (14-17)	(▲)*$_{2/2}$
Said (2022)	MI, mix	E	🟢	Adolescents (14-17)	Adolescents (14-17)	(▲)*$_{2/2}$
Saraf (2015)	MI, rural	E	🟢	Children (6-13)	Children (6-13)	(▲)*$_{3/5}$
Stone (2020)	Mix, mix	SR	🟢	Infants (0-5) Children (6-13)	Children (6-13)	(▲)$_{1}$
von Grafenstein (2023)	MI, N/S	E	🔴	Children (6-13)	Children (6-13)	(▼)$_{4/7}$
Wall (2022)	Mix, rural	SR	🟢	Children (6-13)	Children (6-13)	(▲)*$_{3/3}$
Zhao (2024)	UI, urban	QE	🟡	Children (6-13) Adolescents (14-17)	Children (6-13)	(▲)*$_{1/1}$
Zhou (2016)	UI, N/S	E	🟡	Children (6-13) Adolescents (14-17)	Children (6-13) Adolescents (14-17)	(▲)*$_{2/3}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: 🔴 low quality; 🟡 medium quality; 🟢 high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 28.

SDG 4 – Multiple vulnerability-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Jamaluddine (2022)	MI, rural	QE	●	Children (6-13)	Children (6-13)	(▲)*$_{1/3}$
Jayawardana (2021)	UI, mix	QE	●	Children (6-13) Adults (35-64)	Children (6-13)	(◀▶)$_{1}$
Jeong (2022)	LI, mix	SR	●	HHs receiving humanitarian assistance	HHs receiving humanitarian assistance	(▲)*$_{1/1}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 29.

SDG 5 - Gender-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Austrian (2020)	MI, mix	E	●	Adolescent girls (13-17)	Adolescent girls (13-17)	(▼)*$_{1/1}$
Chinen (2017)	Mix, Mix	SR	●	Young adults (18-34), esp. women	Young adults (18-34), esp. women	(▲)*$_{3/4}$
Gilligan (2015)	UI, urban	E	●	Young adults (18-34)	Young adults (18-34)	(▲)*$_{3/5}$
Gram (2019)	MI, rural	E	●	Mothers or female caregivers (pregnant women)	Mothers or female caregivers (pregnant women)	(▼)$_{3/4}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Peterman (2017)	Mix, N/S	SR	●	Mothers/female caregivers Adolescents (14-17) Young adults (18-34) Adults (35-64)	Focus on children and adolescents	(▲)*$_{1/1}$
Psaki (2022)	Mix, mix	SR	●	Children (6-13) Adolescents (14-17), especially female	Children (6-13) Adolescents (14-17)	(▲)*$_{3/18}$
Spangaro (2021)	LI, N/S	SR	●	Adolescent girls (14-17) and their mothers/caregivers	Adolescent girls and their mothers or caregivers	(▲)*$_{2/4}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects);*(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 30.

SDG 5 - Age-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Peterman (2017)	Mix, N/S	SR	●	Mothers/female caregivers Adolescents (14-17) Young adults (18-34) Adults (35-64)	Focus on children and adolescents	(▲)*$_{1/1}$
Psaki (2022)	Mix, mix	SR	●	Children (6-13) Adolescents (14-17), especially female	Children (6-13) Adolescents (14-17)	(▲)*$_{3/18}$

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Rose-Clarke (2019)	Mix, N/S	SR	●	Adolescents (14-17)	Adolescents (14-17)	(▼)*$_{1/2}$
Roy (2019)	MI, rural	E	●	Mothers or female caregivers Infants (0-5)	Mothers or female caregivers	(▲)$_{5}$
Stark (2018)	LI, rural	E	●	Children (6-13), girls	Children (6-13), girls	(▲)$_{1}$
Wang (2022b)	UI, rural	E	●	Children (6-13) with disabilities (myopia)	Children (6-13) with disabilities (myopia)	(▲)$_{2/2}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

Table 31.

SDG 5 – Multiple vulnerability-responsive

Short name	Context	Design	Evidence strength	Intended beneficiary	Group measure	Direction
Jeong (2022)	LI, mix	SR	●	HHs receiving humanitarian assistance	Women	(▲)*$_{1/1}$
Lwamba (2022)	Mix, mix	SR	●	Adults (35-64)	Women	(◀▶)$_{28}$

Notes: Blue rows refer to in-kind interventions, green rows to voucher interventions, and orange rows to a mix. The following abbreviations are considered for each column.

Context: LI (low-income); MI (lower-middle income); UI (upper-middle income). N/S refers to information not specified.

Design: QE (Quasi-experimental); E (Experimental); SR (Systematic Review)

Strength of evidence: ● low quality; ● medium quality; ● high quality

Direction: ▲ (favourable effect); ▼ (unfavourable effect); ◀▶ (mixed effects); *(statistically significant effect at 1%, 5%, or 10% level); subscripts refer to the share of outcome indicators that align with the specific effects out of the total number of outcome indicators for the specific SDG that were reported.

Source: Synthesis team elaboration

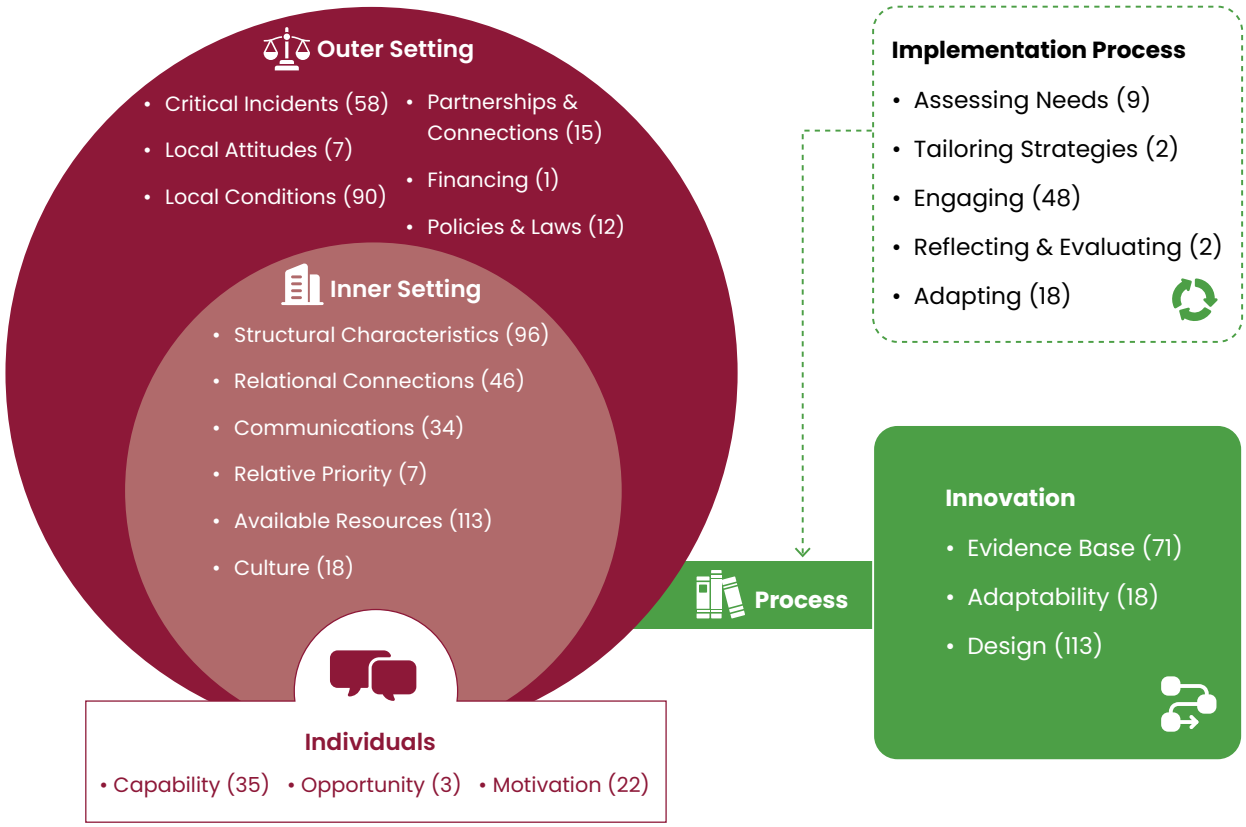
APPENDIX G: Barriers & Drivers – UN evaluations

This appendix presents an overview on the barriers and drivers emerging from 94 UN-led process and performance evaluations. The analysis was guided by a revised version of the Consolidated Framework for Implementation Research, as elaborated by Damschroder et al. (2022). This framework consists of a system of five "domains" (broader categories) and multiple "constructs" (subcategories within domains) which allows for a systematic approach to the identification of factors.

Figure 10 presents an overview on the factors (barriers and drivers) and how many times (n) they emerged from the 94 UN-led process and performance evaluations that were taken into consideration by the qualitative analysis.

Figure 10.

Barriers & Drivers - overview



Source: Synthesis team elaboration, adapted from Damschroder et al. (2022)

The synthesis team subsequently used the framework of the OECD DAC criteria to structure the findings by determining whether they affect "coherence," "relevance," "effectiveness," or "sustainability." A summary depiction of these findings is presented in Tables 32-35.

Table 32.

Coherence - drivers & barriers

Domain (n)	Construct Name (n)	Drivers/ Barriers (n)	Examples
Inner Setting (20)	Structural Characteristics (13)	0/13	• Absence of an overarching theory of change at Country Office (Luzot 2023c) • Available funding, differences in budget hinder collaboration under "Delivering as One" approach (Jordan 2021) • Competition for resources hinders collaboration (Faúndez 2022b)
	Available Resources (5)	0/5	
	Culture (2)	0/2	
Outer Setting (9)	Partnerships & Connections (6)	2/4	• Lack of technical forum bringing together UN agencies and implementing partners (Knips 2022) • "Delivering as One" approach as UN-wide initiative promotes coherence Crises serving as catalyst for coordinated strategies (Faúndez 2022b)
	Policies & Laws (2)	1/1	
	Critical Incidents (1)	0/1	
Individuals (4)	Motivation (4)	3/1	• Coherence of interventions at Country Office as the result of the initiative of individuals initiative (Bester 2022)
Innovation (3)	Innovation Design (3)	1/2	• Issues addressed by the intervention; issues are compatible with challenges other organisations address with their programmes (Bhattacharjee 2023)

Source: Synthesis team elaboration

Table 33.

Relevance - drivers & barriers

Domain (n)	Construct Name (n)	Drivers/ barriers (n)	Examples
Innovation (90)	Innovation Evidence Base (63)	42/21	• Programme design based on intense analysis, resulting in, for instance, a focus on vulnerable groups (Chavez 2022) • Targeting criteria neglect major causes for school drop-out (Ficini 2022) • Standardised approaches applied by donors do not allow for sufficient flexibility to adjust to local needs (Luzot 2023)
	Innovation Design (14)	4/10	
	Adaptability (13)	8/5	
Inner Setting (52)	Structural Characteristics (18)	11/7	• UNICEF's presence across the country conducive to partnerships with local partners and the capability to address the most urgent needs (Bhattacharjee 2019) • Unavailability of updated data base (Cano 2022) • Strong trustful connections with the Government (Luzot 2022a) • WFP in continuous dialogue with Government and other stakeholders to ensure that actions stay aligned with national needs (Cook 2022c) • Evolution of governmental priorities led to challenges in staying relevant (Jordan 2021)
	Availability of Resources (10)	2/8	
	Relational Connections (9)	6/3	
	Communications (8)	5/3	
	Relative Priority (7)	0/7	
Implementation Process (28)	Engaging (15)	13/2	• Close consultation of vulnerable groups during implementation (FAO 2023a) • Regular data collection throughout implementation informs programme adaptations (Luzot 2023b) • Inappropriate adaptations (implemented without sufficient needs assessment) (Fuentes 2023)
	Assessing Needs (9)	6/3	
	Adapting (4)	2/2	
Outer Setting (6)	Local Conditions (5)	0/5	• Similar services offered by other actors in the same location (Cano 2022) • Distributions not matching local diet habits (Luzot 2022b)
	Local Attitudes (1)	0/1	
Individuals (1)	Capability (1)	1/0	• Project staff's knowledge and creativity contributed to alignment with the needs of vulnerable groups (Salinas 2022)

Source: Synthesis team elaboration

Table 34.

Effectiveness - barriers & drivers

Domain (n)	Construct Name (n)	Drivers/ Barriers (n)	Examples
Inner Setting (165)	Availability of Resources (67)	2/65	• Limited availability of funds restricted scope of interventions (Noiji 2024) • High rates of staff turnover (Luzot 2022a) • Long, trustful relations with partners (Cruz 2021a) • of sense of accountability (Peacocke 2018) • Good internal communication (Cruz 2021b)
	Structural Characteristics (49)	11/38 26/6	
	Relational Connections (32)	4/6	
	Culture (10)	4/3	
	Communications (7)		
Outer Setting (123)	Local Conditions (62)	1/61	• Lack of infrastructure (Pompili 2023b) • Closure of schools due to the Covid-19 pandemic (Larmoter 2023) • UNICEF's ability to build on pre-existing (prior to the intervention) partnerships with CSO when partners needed to be replaced (Bhattacharjee 2023) • Hostility towards migrant population affected the provision of assistance (Cook 2022a)
	Critical Incidents (52)	2/50	
	Partnerships & Connections (6)	3/3	
	Local Attitudes (3)	0/3	
Innovation (46)	Innovation Design (41)	13/28	• Sound design of the intervention (Cook 2021b) • Lack of analysis led design to be based on wrong assumptions (IPEC 2016)
	Innovation Evidence Base (5)	2/3	
Individuals (46)	Capability (27)	9/18	• Poor skills of caretakers (Peacocke 2018) • High dedication of WFP project staff (Cook 2022c) • School children need to work instead of attending schools (Mailloux 2019)
	Motivation (16)	6/10	
	Opportunity (3)	0/3	
Implementation Process (29)	Adapting (14)	11/3	• WFP's ability to adapt the intervention to pandemic-related restrictions (Cook 2022c) • Implementing partners selected based on administrative criteria instead of technical skills (FAO 2023c) • Design short of measures to make intervention gender-/disability-responsive but during implementation, activities were undertaken to ensure inclusivity (Cano 2022) • Insufficient monitoring prevented effective adjustments (Godden 2017)
	Engaging (11)	7/4	
	Tailoring Strategies (2)	2/0	
	Reflecting & Evaluating (2)	0/2	

Source: Synthesis team elaboration

Table 35.

Sustainability - barriers & drivers

Domain (n)	Construct Name (n)	Drivers/ Barriers (n)	Examples
Innovation (63)	Innovation Design (55)	30/25	• Lack of exit strategies (Muñoz 2021) • Universal nature of stipend granted for school meals (Peacocke 2022) • Intervention based on insufficient understanding of the target group's economic strategies (Cook 2021a)
	Adaptability (5)	2/3	
	Innovation Evidence Base (3)	1/2	
Inner Setting (60)	Available Resources (31)	5/26	• Financial constraints in government ministries (Bester 2022) • Weak governmental capacities (Chavez (2022) • Lack of governmental commitment to take ownership (Odimengwu 2021) • Close collaboration between communities and government (Das 2020) • Failure to communicate relevant information to beneficiaries (Das 2020)
	Structural Characteristics (16)	3/13	
	Culture (6)	4/2	
	Relational Connections (5)	2/3	
	Communications (2)	0/2	
Outer Setting (44)	Local Conditions (23)	0/23	• Intervention implemented in a fragile context (FAO 2023c) • Absence of legal framework regulating school feeding (Ficini (2022) • Elections and related restructuring of political landscape (Bockel 2022) • Private sector engagement (Byabagambi 2021) • Persistence of detrimental practices like early marriages, even among persons who were meant to be catalysts of change (Fuentes 2023)
	Policies & Laws (10)	2/8	
	Critical Incidents (5)	0/5	
	Partnerships & Connections (3)	2/1	
	Local Attitudes (3)	1/2	
	Financing (1)	0/1	
Implementation Process (22)	Engaging (22)	16/6	• Engagement of town halls and local elected representatives (Ficini 2022)
Individuals (9)	Capability (7)	2/5	• Insufficient skills among beneficiaries to keep activities going (Turano 2022) • Level of satisfaction among beneficiaries (Peacocke 2022)
	Motivation (2)	0/2	

Source: Synthesis team elaboration

APPENDIX H: SYSTEM-LEVEL OUTCOMES AND PROGRAM CONTRIBUTIONS

This appendix presents evidence from 94 UN-led process and performance evaluations on social protection system-level outcomes (which could be directly or indirectly linked to gender-, age-, disability- responsive vouchers and in-kind transfers) as well as programme outputs which contributed to the changes The tables below map system level outcomes and programme contributions, using a framework which combines common UN building blocks for social protection systems, and synthesise evidence on contributing factors to system-level changes beyond programme outputs by using the Consolidated Framework for Implementation Research (Damschroder et al. 2022) to frame the analysis.

Table 36.

System-level outcomes and program contributions

Building Block	Definition	System-Level Outcomes (No. of studies)	Programme Contributions	Examples
1. Legislation	Laws and legal frameworks that establish social protection entitlements and programme authority	• Adoption of new social protection legislation (12) • Revision of existing social protection laws (3)	• Technical assistance for drafts / revisions • Advocacy • Evidence generation • Capacity development	"Through specialised technical assistance, it collaborated with FAO and the Ministry of Education in the drafting of the Law School Feeding Organic Law." (Cook 2022c)

2. Policy, Plans & Strategies	National strategies, policies, and operational plans guiding social protection programmes	• Policy / plan / strategy formulation (28) • Policy / plan / strategy adjustments (16) • Formulation of policies / plans / strategies with multisectoral approach (4)	• Technical assistance for drafts / revisions • Advocacy • Evidence generation • Capacity development • Financing support • Drafting plans / strategies • Good practices from (pilot) programmes scaled up	"WFP was able to successfully make the case for mainstreaming fortified rice in government safety nets in 2019. This translated into the development of a Centre Sector Scheme; WFP supported both the formulation of the scheme as well as the development of the operational guidelines for its cascade down to the states." (Luzot 2022a)
3. Law / Policy Implementation - Service Delivery	Operationalization of laws and policies through actual social assistance programme delivery	• Improved institutional capacities incl. standards, structures, mechanisms and operating procedures (29) • Increased or improved coverage (8) • Improved financing and resource allocation (6) • Continuation of service delivery during emergencies (2) • increasing government ownership of provided service (2)	• Technical assistance incl. provision / co-development of tools & guidelines • Capacity development incl. provision of training • Advocacy • Evidence generation • Provision/ facilitation of funding	"WFP also contributed to increasing the coverage of the school feeding programmes implemented by the Ministry of Education and the Ministry of Health and in 2021 a budget line for school feeding was included in national and state budgets for the first time." (Luzot 2022b)

4. Evidence & Data Systems	Governmental data collection, monitoring, evaluation, and knowledge management practices and systems related to social assistance	• Improved data collection and management systems (8) • Improved / increased evidence generation (6)	• Evidence generation • Technical assistance • Capacity development incl. training • Provision of resources (IT infrastructure)	"Strengthening the nation's social protection framework capacity building through training on a blockchain-based system to include it in the national humanitarian aid system […] Rahat complements the government's social protection schemes by enhancing targeting precision, minimising resource leakage, and enabling real-time monitoring, thereby amplifying the effectiveness of these initiatives" (Basnet 2023)

Source: Synthesis team elaboration

Table 37.

Factors influencing system-level change

Construct Name	Evaluations & Factors (N)	Inner Setting Location (N)	Description (reference)
Inner Setting Domain			
Structural characteristics: Work Infrastructure	Evaluations (8) Drivers (2) Barriers (10)	Govt. (9) UN (3)	• Training design/frequency not compatible with high staff turnover (Arzeni 2020) • Change of government and high turnover of (administrative and political) staff undermines the Institutionalization of programmes and services (Cook 2022a) • Existing (strong) institutional capacities and platforms for multisector coordination (Arzeni 2020) • Strong existing institutions and legal frameworks to sustain programme effects (UNICEF 2023c) • Limited institutional capacity undermines policy / programme implementation or expansion (Bahr 2022, Bhattacharjee 2023, Pompili 2023b) • Limited capacity of government in procurement and storage of food (Luzot 2023b) • Limited capacity of government to provide social assistance services in areas under the control of armed groups (Pompili 2023b) • Difficulty in util ising WFP's corporate outcome and output indicators as a basis for measuring programme progress (Bonino 2023) • Short-term support through WPF not sufficient for institutionalization, and for support to be translated into public policy (Cook 2022a) • Lack of consolidated corporate WFP framework for institutional capacity building as missed opportunity to contribute to programme efforts (Pompili 2023b)
Structural characteristics: Physical Infrastructure	Evaluations (2) Barriers (2)	Govt. (2)	• Existing infrastructure (incl. equipment) at schools insufficient for optimal roll out of national school feeding law (Knips 2022). • Limited capacity of government in procurement and storage of food (Luzot 2023b)

Construct Name	Evaluations & Factors (N)	Inner Setting Location (N)	Description (reference)
Relational connection & communication	Evaluations (5) Drivers (4) Barriers (1)	Govt. (2) UN (3)	• Existing (strong) institutional govt. capacities and platforms for multisector coordination were further strengthened (Arzeni 2020) • Productive partnership World Bank has enabled UN to significantly strengthen government capacities in social protection (Jordan UN 2021) • UNICEF's strong track record in coordination and reputation as the "go-to" agency for system strengthening as an opportunity and added value towards joint (social protection) system-strengthening efforts (Kacapor-Dzihic 2022) • Creation of donor alliances and partnerships leads to establishment of social protection policy as reference framework for funding of social protection services (Luzot 2023) • Coordination with other actors (incl. implementers) flagged as challenge that needs to be overcome to ensure alignment and coordinated work (Women 2023)
Culture or relative priority	Evaluations (4) Drivers (2) Barriers (1) Unclear (1)	Govt. (4)	• Sustainability of outputs/outcomes dependent on political will and priorities of coming decision makers (Arzeni 2020) • Strong ownership /prioritization towards policy or programme implementation by government (Bhattacharjee 2023, Luzot 2023) • Advocacy & evidence changed government's vision towards improved social assistance service delivery (Cook 2022b) • Lack of ownerships / prioritization of government to mainstream gender (into social protection schemes) (Jordan UN 2021)
Available resources: funding	Evaluations (5) Driver (1) Barriers (4)	Govt. (4) UN (1)	• Lack of financial resources to sustain programme activities (Arzeni 2020) • Funding shortages / limited financial resources limit program/ policy implementation (Bahr 2022; Bhattacharjee 2023, Pompili 2023b) • Creation of donor alliances and partnerships leads to establishment of social protection policy as reference framework for funding of social protection services (Luzot 2023)

Construct Name	Evaluations & Factors (N)	Inner Setting Location (N)	Description (reference)
	Evaluations (5) Drivers (2) Barriers (3)	Govt. (3) UN (2)	• Strong existing institutions and legal frameworks to sustain programme effects (UNICEF 2023b) • Training design/frequency not compatible with high staff turnover (Arzeni 2020) • Short- term support through WPF not sufficient for institutionalization, and for support to be translated into public policy (Cook 2022a) • Difficulty of programme to utilise WFP's corporate outcome and output indicators as a basis for measuring progress (Bonino 2023) • UNICEF's strong track record in coordination and reputation as the "go-to" agency for system strengthening as an opportunity and added value towards joint (social protection) system-strengthening efforts (Kacapor-Dzihic 2022)
Outer Setting Domain			
Local conditions	4	1/3	• Adverse local conditions (especially in emergencies) restrict beneficiaries' access to non-digital social assistance distribution (Basnet 2023) • Field access constraint of WFP, likely due to adverse local conditions (Peacocke 2022) • Reach of the government (to implement social assistance) limited as much territory in hands of armed groups (Pompili 2023b) • Complementary food assistance activities to establish favourable local market conditions (Cook 2022b)
Critical incidents	4	0/4	• Adverse local conditions (especially in emergencies) restrict beneficiaries' access to non-digital social assistance distribution (Basnet 2023) • Potentially unexpected change of government and high (technical & political) staff turnover undermines the institutionalization of programmes and services (Cook 2022a) • Despite political instability (assassination of president and subsequent political turmoil) technical assistance established to implement social protection policy (Luzot 2023) • COVID 19 as barrier to institutionalization of programmes / services (Cook 2022a)

Construct Name	Evaluations & Factors (N)	Inner Setting Location (N)	Description (reference)
Local attitudes	3	2/1	• Lack of ownerships / prioritization of government to mainstream gender (into social protection schemes) (may also reflect broader local attitudes) • Programme strengthens capacity of affected people to hold government accountable by raising awareness about their entitlements (Luzot 2022) • Transformation of cultural codes for the transformation of social problems (Chavez 2022)
Partnership & connections	2	1/1	• Partnership and dialogue accelerates social legislation and policy (through working group) a (Bhattacharjee 2023) • Envisaged system-change blocked by parts of the government (Bockel 2023)
Laws & policies	1	1/0	• Strong legal frameworks facilitate system-strengthening efforts (UNICEF 2023c)
Implementation process domain			
Assessing needs / context	4	3/1	• Programme beneficiary needs / situation data and analysis enhances public knowledge and feed into policy development (Cabo 2023; Cook 2023c) • Assessment of existing service delivery feeds into policy implementation (Bajraktari 2023) • Inconsistent capacity strengthening due to lack of investment in understanding the capacity support needs of key partners (programme deliverers) (Cook 2021)
Teaming	2	2/0	• Productive partnership with World Bank has enabled UN Jordan to significantly strengthen government capacities in social protection (Jordan UN 2021) • UNICEF's good reputation in system strengthening makes other stakeholders want to team up for joint programming in this field (Kacapor-Dzihic 2022)
Planning/ tailoring strategies	1	0/1	• Smaller scope of capacity development efforts/ outcomes than expected as activities not structured on the basis of a national capacity strengthening plan (Cook 2022a)

Construct Name	Evaluations & Factors (N)	Inner Setting Location (N)	Description (reference)
Innovation domain			
Innovation design	2	1/1	• "Technology based system" important for more effective service delivery (social registry & monitoring) (Basnet 2023) • Limited effects on system strengthening due to "inadequacy of the consultancy services provided" (Pompili 2023b)
Innovation: trialability, evidence base & adaptability	2	2/0	• Public regulatory framework adopted that allows the scaling up of the models tested by the WFP on a small scale (Cook 2022b) • Law on School Feeding and its regulations drafted based on the successful results of the school feeding supplementation project (Cook 2022c)
Individuals domain			
Implementation Facilitators: capability	2	0/2	• Insufficient experience & expertise of UN country office staff in the field of system strengthening (Cook 2021) • Limited effects on system strengthening due to "inadequacy of the consultancy services provided" (Pompili 2023b)

Source: Synthesis team elaboration

The Global SDG Synthesis Coalition

The Global SDG Synthesis Coalition brings together evaluation offices from the United Nations, multilateral development banks, and international financial institutions to generate and share evidence on what works, how, and why to advance the Sustainable Development Goals (SDGs). Co-chaired by the *Independent Evaluation Office of UNDP* and the *Independent Evaluation Department of the Asian Development Bank*, the Coalition promotes joint learning and actionable syntheses that inform policy and practice toward achieving the 2030 Agenda.


sdgsynthesiscoalition.org


SDGSynthesis


sdgsevidence


The Global SDG Synthesis Coalition