



CEDIL methods working paper: Making and justifying evidence claims: evidence synthesis of impact evaluations and systematic reviews in international development

About CEDIL

The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK government through UK Aid. The mission of the Centre is to test innovative methodologies in evaluation and evidence synthesis and promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

CEDIL syntheses working papers

The CEDIL Syntheses Working Paper series publishes innovative research syntheses of the impact of interventions in low- and middle-income countries in the form of evidence and gap maps, systematic reviews, meta-analyses and other synthesis products.

Suggested citation: Bangpan, M., Draper, A., Minchenko, D., Stansfield, C., Dickson, K., Tripney, J. and Oliver, S. (2023). *CEDIL methods working paper: Making and justifying evidence claims: evidence synthesis of impact evaluations and systematic reviews in international development, CEDIL Syntheses Working Paper 10, CEDIL, Oxford*. Available at <https://doi.org/10.51744/CSWP10>

This project was funded by the Centre of Excellence for Development Impact and Learning (CEDIL), supported by UK Aid from the UK Government. The views expressed in this working paper do not necessarily reflect the UK Government's official policies.

Cover design: PhilDoesDesign

Cover image credit: Thamizhparithi Maari, shared on flickr under a Creative Commons licence

Copyright: © 2023 This is an open-access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

CEDIL methods working paper: Making and justifying evidence claims: evidence synthesis of impact evaluations and systematic reviews in international development

Authors

Mukdarut Bangpan

Aisling Draper

Dayana Minchenko

Claire Stansfield

Kelly Dickson

Janice Tripney

Sandy Oliver

Affiliation

The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)

Social Research Institute

University College London



Contents

List of abbreviations	7
Executive summary	8
1 Background.....	11
1.1 Evidence claims for policy decisions.....	11
1.2 Standards, values and framing of evidence claims.....	11
1.3 Evidence claims for global South decisions.....	13
1.4 Frameworks for evidence claims.....	14
1.5 Research questions	17
2 Methods	18
2.1 Identifying studies	18
2.2 Coding and data extraction.....	22
2.3 Analysis approaches.....	23
3 Making and justifying claims in international development studies.....	25
3.1 Focus of the claims.....	25
3.1.1 Empirical knowledge claims of impact evaluations	25
3.1.2 Methodological claims of impact evaluations.....	29
3.1.3 Empirical knowledge claims of systematic reviews	32
3.1.4 Methodological claims of systematic reviews.....	36
3.2 Scope of claims.....	39
3.2.1 Generalisability and transferability in impact evaluations	39
3.2.2 Generalisability and transferability in systematic reviews	40
3.3 Communicating evidence claims.....	41
3.4 Framing of reviews to make evidence claims, with social values for impact	42
3.4.1 Framing of reviews.....	42
3.4.2 Social values shaping reviews	43
3.4.3 Reviews with impact.....	44
4 Other literature relevant to making and justifying evidence claims in research...48	
4.1 Descriptions of relevant literature making and justifying claims.....	48
4.2 Summary of overarching literature on justifying claims (n=9)	50
4.3 Summary of literature focusing on particular aspects of justifying evidence claims (n=36).....	51

4.3.1	Research design and making a claim.....	52
4.3.2	Triangulation	53
4.3.3	Validity and interpretation of research findings for making a claim...54	
4.3.4	Relevance and usefulness of making claims.....	55
4.3.5	Communication of claims	56
4.3.6	Other papers not considered (n=34).....	57
5	Discussion	58
5.1	Summary of findings	58
5.1.1	Focus and scope of the claims in impact evaluations	58
5.1.2	Focus and scope of the claims in systematic reviews	61
5.1.3	Framing of systematic reviews and their underpinning social values	64
5.2	Strengths and limitations	64
5.3	Conclusion and recommendations.....	66
ANNEX A:	Fitness for purpose of an evidence claim framework (Gough, 2021).....	69
ANNEX B:	Data extraction tools	71
ANNEX C:	Findings: a review of impact evaluations and systematic reviews.....	84
	Key characteristics of the impact evaluations and systematic reviews included in the review.....	84
	Key characteristics of impact evaluations.....	86
	Key characteristics of systematic reviews.....	90
ANNEX D:	Key characteristics of studies	1
ANNEX E:	Social values in health and care guidance	43
ANNEX F:	Framing of systematic reviews to make evidence claims	45
Annex G:	Truth table of attributes for developing evidence claims	50
References	2

List of figures

Figure 1: Framework for developing and evaluating complex interventions (Skivington et al. 2021).....	115
Figure 2: Framework for developing and evaluating research methods	116
Figure 3: Framework for developing, justifying, and communicating evidence claims	117
Figure 4: Overall approach of the review	118
Figure 5: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram	85
Figure 6: Type of policy sectors covered by impact evaluations, by region (codes not mutually exclusive)	88
Figure 7: Study design (codes not mutually exclusive)	90
Figure 8: Systematic reviews and type of policy sectors by income level and geographical focus (codes not mutually exclusive).....	91

List of abbreviations

CASP	Critical Appraisal Skills Programme
cRCT	Cluster randomised controlled trial
FCDO	UK Foreign, Commonwealth and Development Office
GRADE	Grading of Recommendations Assessment, Development and Evaluation
LMIC	Low- and middle-income countries
MMAT	Mixed Methods Appraisal Tool
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	Randomised controlled trial
SDGs	Sustainable Development Goals
SES	Socio-economic status

Executive summary

In the last few decades, there has been an increasing use of, and demand for, high-quality evidence among decision makers in international development. Evidence from research leads to claims about the effects of intervening in people's lives to improve specific outcomes. However, this raises the question of how such evidence claims are framed, justified and communicated. This review aims to address this important question by reviewing research findings from impact evaluations and systematic reviews to understand the nature and the scope of evidence claims in the global South.

Overall, we found that most of the evidence claims in the studies we reviewed are framed by research teams from medium- or high-income countries (n=75), with fewer than 5% of included studies and reviews led by researchers from low-income countries (n=3). Of the included studies and reviews, the nature of the claims focus broadly on intervention outcomes and effects (impact), implementation factors, the choice of research designs and methods, the credibility of the research, and knowledge gaps. Claims and justifications are typically communicated using a variety of tables, graphs, diagrams, and geographical maps, while claims about empirical knowledge are communicated using standardised outputs, such as forest plots, charts and online evidence maps (a visual presentation of the body of evidence).

Within the **impact evaluations** included in our review, we found that claims about intervention impacts are primarily justified in terms of their technical quality, namely their certainty, strength and consistency. This justification is occasionally complemented by consideration of the appropriateness of the study design and research methods to address the research questions.

Claims about methodology are justified by considering whether the findings warrant conclusions regarding causality. Meanwhile, claims about quality and validity are implicitly justified through the employment of well-established study designs (e.g. randomised controlled trials (RCTs)) or triangulating data. When studies are not RCTs, claims about quality and validity are often justified by reference to their feasibility, including whether the research design is appropriate to address review questions or for research in humanitarian contexts.

For claims about generalisability, the representativeness of the study population, the transferability of the findings to multiple settings, contextual similarity, and the programme theory underpinning research and implementation are often used to justify such claims in the impact evaluations included in the review.

Within the systematic reviews, we found that evidence claims about intervention impacts are justified based on four aspects related to the evidence included in the reviews: quantity, quality, consistency and the degree of coherence of the focus. Where findings are inconclusive or no evidence of impact is reported, inconsistency or a lack of evidence are often used to justify such claims. Research gaps are warranted when evidence is insufficient, judged to be of poor quality, or inconsistent.

Review approaches used in the systematic reviews are justified according to the availability and appropriateness of study designs to answer the review questions. The consistency of evidence and the perspectives of reviewers and stakeholders are also used to justify chosen approaches. Similarly, claims made about the quality and validity of review findings are justified by employing accepted methodological standards and procedures to address the review questions.

Social values are apparent in all of the systematic reviews included in our review. Neither the process of framing (either from the literature and/or with key stakeholders) nor the structure of framing (for example, the ecological model, theory of change or logic model) are associated with specific review methods. However, involving stakeholders in framing the review and developing a theory of change, logic model or ecological model may have helped to identify the social values that contributed to framing the reviews.

In regard to generalisability, applicability and implementation challenges are discussed in systematic reviews, rather than making claims about generalising findings to other new contexts.

Research implications

Claims within research often go beyond the issue of impact, and thus claims about equity, innovation, scale-up and sustainability, and how they are framed and justified, should be further explored. This can inform the development of appropriate and novel research designs and methods that can be used to produce credible, reliable and relevant evidence to support such claims confidently.

Ethical considerations should also be routinely considered when designing impact evaluations and implementing programmes or policies to enhance applicability, whilst producing socially responsible research.

Diversity of evidence can play an important role in supporting claims about sustainability and scale-up. Bringing evidence from various research designs, sources,

voices, and origins strengthens collaboration and the pertinence to evidence-informed decisions in international development. Having clear stakeholder engagement strategies, understanding local and citizen knowledge systems, building research capacity, and creating more equal partnerships between those involved in the co-production of evidence may increase the likelihood of long-term programme impact and sustainability.

How did we get the results?

We included within our review a total of 78 studies (47 impact evaluations and 31 systematic reviews). We searched 3ie's Development Evidence Portal and the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) website. Impact evaluations and systematic reviews were included if they were funded by the UK Foreign, Commonwealth, and Development Office (FCDO) (formerly the Department for International Development), published in 2017 or later, and assessed the impact of any intervention programme in low- and middle-income countries (LMICs). Additional supplementary searching was carried out to ensure that studies and reviews focusing on less common policy sectors were also included in the review. We extracted data to identify claims about empirical findings, knowledge gaps, methodology and reach. Once extracted, the findings were synthesised using a framework synthesis. Additional literature from other disciplines was narratively described, summarising the current debates related to making and justifying claims.

1 Background

1.1 Evidence claims for policy decisions

The Centre of Excellence for Development Impact and Learning (CEDIL) aims to develop and test innovative methods in evidence synthesis and impact evaluations relevant to low- and middle-income economies, and to build a greater evidence base on the use of research for decision-making. In the last few decades there has been increasing demand to make better use of the findings of impact evaluations and evidence synthesis. The findings of such studies have led to claims about the effects of intervening in people's lives that have influenced decisions about programmes or policies. The Cambridge Dictionary defines a claim as '*a statement that something is true or is a fact, although other people might not believe it*'.¹ In this context, we define an evidence claim as a statement that is said to be supported by research evidence (or it is implied that this is the case), even though other people may question whether the evidence supporting that statement has been compiled or scrutinised appropriately.

This raises the important question of how such 'evidence claims' from research are compiled or scrutinised appropriately – is the research underlying the claims valid, appropriate, or relevant to decision-making, and how is this judged? Scrutinising evidence claims is a core part of the methodology for synthesising research and for making appraised evidence readily available for policymakers and practitioners. However, a survey of 14 publicly available English language evidence portals in Europe and the USA found that the criteria for appraising evidence (or 'evidence standards') vary across portals, though they focus predominantly on questions about the efficacy of interventions that are answered by conducting experimentally controlled studies (Gough and White, 2018).

1.2 Standards, values and framing of evidence claims

Extensive research has been undertaken with the aim of developing checklists and reporting standards for assessing the quality of research studies. These include checklists and reporting standards for trials (Moher *et al.*, 2012), observational studies (von Elm *et al.*, 2007), realist evaluations (Wong *et al.*, 2016), economic evaluations (Husereau *et al.*, 2022), implementation studies (Pinnock *et al.*, 2017), qualitative research (O'Brien *et al.*, 2014; Tong *et al.*, 2007), systematic reviews (Page *et al.*, 2021;

¹ <https://dictionary.cambridge.org/dictionary/english/claim>

Stroup *et al.*, 2000; Wong *et al.*, 2013; Higgins *et al.*, 2018); for assessing the quality of research (Shea *et al.*, 2017; Sterne *et al.*, 2016; Whiting *et al.*, 2016; Bero *et al.*, 2018; Sterne *et al.*, 2019); and for developing evidence-informed recommendations (Brouwers *et al.*, 2010). However, these tools focus largely on the methodological and reporting quality of evidence, with less consideration given to the values and framings underpinning claims, possible alternative values and framings, and the clarity and relevance of claims for decision-making. Gough (2021) identified this limitation and proposed the Fitness for Purpose of an Evidence Claim Framework, which introduces core issues that are relevant to appraising the ‘fitness for purpose’ of evidence claims.

The Fitness for Purpose of an Evidence Claim Framework first discusses the nature of evidence claims, which can vary depending on the perspectives of evidence producers and users, and how they are framed in terms of the scope (breadth and detail), level of analysis,² focus, certainty (precision) and generalisability of the claim. The framework also considers whether evidence claims can be warranted, or justified, by the appropriateness, quality and relevance of the research methods employed and the resulting findings. Next, the framework refers to different types of evidence standards, guidelines and tools for producing, reporting and assessing the quality of research evidence, and how these tools and guidance are applied to make and justify evidence claims (Gough, 2021). Finally, the framework examines approaches and mechanisms through which evidence claims might be established and communicated to users.

The Fitness for Purpose of an Evidence Claim Framework was subsequently applied to appraise evidence claims made by studies on basic income, to evaluate the extent to which the claims were trustworthy and fit for purpose in addressing the review questions (Chrisp *et al.*, 2022).

Another limitation of checklists and standards, and of Gough’s critique and broader conceptual work, is their emphasis on evidence claims in the global North, where impact evaluations and evidence syntheses are plentiful. Here, we turn our attention to claims about socio-economic development in the global South.

² For this review, ‘level’ refers to the different aspects of analysis where evidence claims are made: for example, at the population, community, village or individual level.

1.3 Evidence claims for global South decisions

In preparation for the current study, we explored debates and practices among the authors of CEDIL-commissioned studies (Oliver *et al.*, 2023). We describe that work briefly here.

We compiled and analysed the following: guidance from CEDIL; discussions at a workshop convened by CEDIL to discuss evidence to enable policy, practice and decision-making; and outputs from empirical studies that form CEDIL's research programme about evaluating complex interventions, enhancing evidence transferability, and increasing evidence use. The workshop discussion encouraged focusing research on the following: the demand for evidence to design research that strengthens confidence in the evidence claims that matter most for decision-making; promoting the uptake of methodological innovations through professional evaluation associations; and wider peer review to ensure greater clarity of synthesis and interpretation.

In a preliminary Lessons Learnt paper, analysis of claims made in CEDIL-commissioned studies led to tentative findings about the nature of evidence claims, and how they were justified and communicated (Oliver *et al.*, 2023). The findings are summarised briefly here, before we set out to apply them to a wider literature:

- **Nature of claims:** Methods for assessing evidence claims originally focus on claims about the effects of an intervention in areas where synthesised evidence is plentiful. By focusing an analysis of evidence claims on a set of studies from the global South, discussion about the nature of claims has broadened in several dimensions. Research reports commissioned by CEDIL make both methodological claims and empirical claims; claims resting on single studies and bodies of evidence; claims about developing and implementing interventions, not only their effects; and claims about the influence of contextual factors. These include empirical claims about the following: knowledge gaps (what we do not know, and what is important to know); and different forms of empirical findings (e.g. effectiveness, the relevance of an intervention, technical feasibility and acceptability, contextual influences, implementation issues, cost-effectiveness analysis). Complementing these are methodological claims (appropriate choice or development of study methods and tools). Empirical and methodological claims can be based on the following: individual studies; summaries of bodies of evidence, research and broader implications; and guidance based on study findings.

Claims can be predictions about what might be found, which is a necessary part of hypothesis testing for empirical studies and is valuable for testing methods claims. Alternatively, claims may be retrospective, about what has been found from observation or exploration, which is an important part of theory building or methods development.

- **Justifying claims:** Some claims are justified by applying methods or tools to assess the validity of methods (and findings) based on prior methodological research or adopting or adapting methods or tools used in another empirical study. The justification for claims may be implicit, particularly when adopting or adapting well-established methods. Earlier methodological or empirical papers may be cited to support the approach, or readers of academic reports may be assumed to share a similar background knowledge. When developing innovative methods, claims may be justified by detailed argument, with or without highlighting uncertainties, with ‘proof of concept’ studies, or by citing other methodological studies.
- **Communicating claims:** The degree of confidence in claims is not always communicated clearly: authors may rely on readers’ understanding of the implications of methodological choices. Where the degree of confidence in the claim is explicit, this may be enhanced for a wide readership by the use of standardised text, scales or scores, or images.

1.4 Frameworks for evidence claims

Empirical claims resting on well-established methodologies tend to be implicit. Their breadth is reflected in an established framework (Figure 1) for developing and evaluating complex interventions for health and social care services, public health practice, and other areas of social and economic policy (Skivington *et al.*, 2021). The first versions of this framework focused on RCTs of complex interventions (Campbell *et al.*, 2000). Later developments accommodated additional formative evaluations, process evaluations, observational designs when experimental designs cannot be applied, and tailoring interventions to local circumstances (Craig *et al.*, 2008). The latest framework’s placing of the context of the evaluation and engaging stakeholders at the core, to be considered at each stage of the work, makes it particularly suitable for framing the development and evaluation of social and economic development interventions.

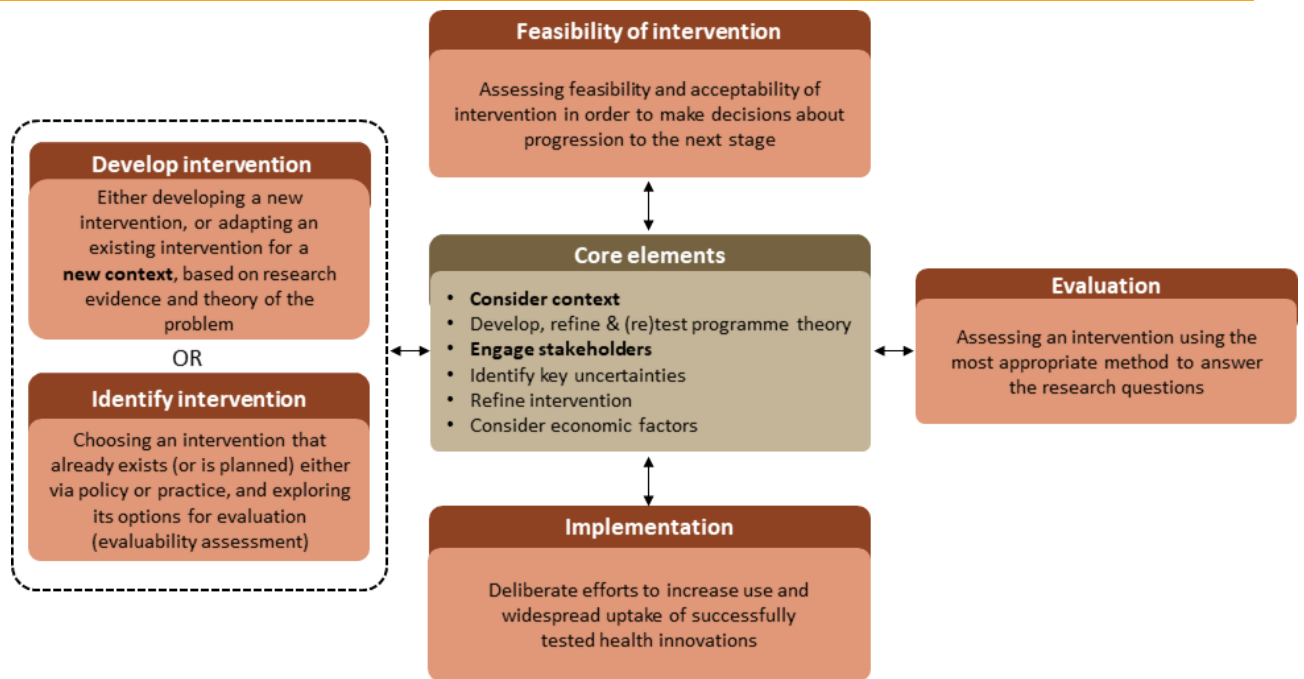


Figure 1: Framework for developing and evaluating complex interventions (Skivington *et al.* 2021)

In addition to empirical claims, CEDIL authors make methodological claims about the choice of methods, about innovative methods for measuring effect sizes, about confidence in causal claims, about taking research findings from one context to use in another, and about innovative combinations of methods. Unlike most empirical claims, claims of methodological innovation are supported more explicitly by detailed argument, with ‘proof of concept’ studies, or by citing other methodological studies. These well-established approaches for developing and evaluating research and evaluation methods can be presented in an analogous conceptual framework (Figure 2).

We postulate that implicit claims resting on well-established methods may well cluster around fields that have attracted academic attention over a longer timescale, commonly in the global North. In contrast, explicit empirical claims, which are more often justified by arguments for methodological innovation, may well cluster around the more challenging policy fields or under-resourced or fragile settings that have a shorter history of academic attention.

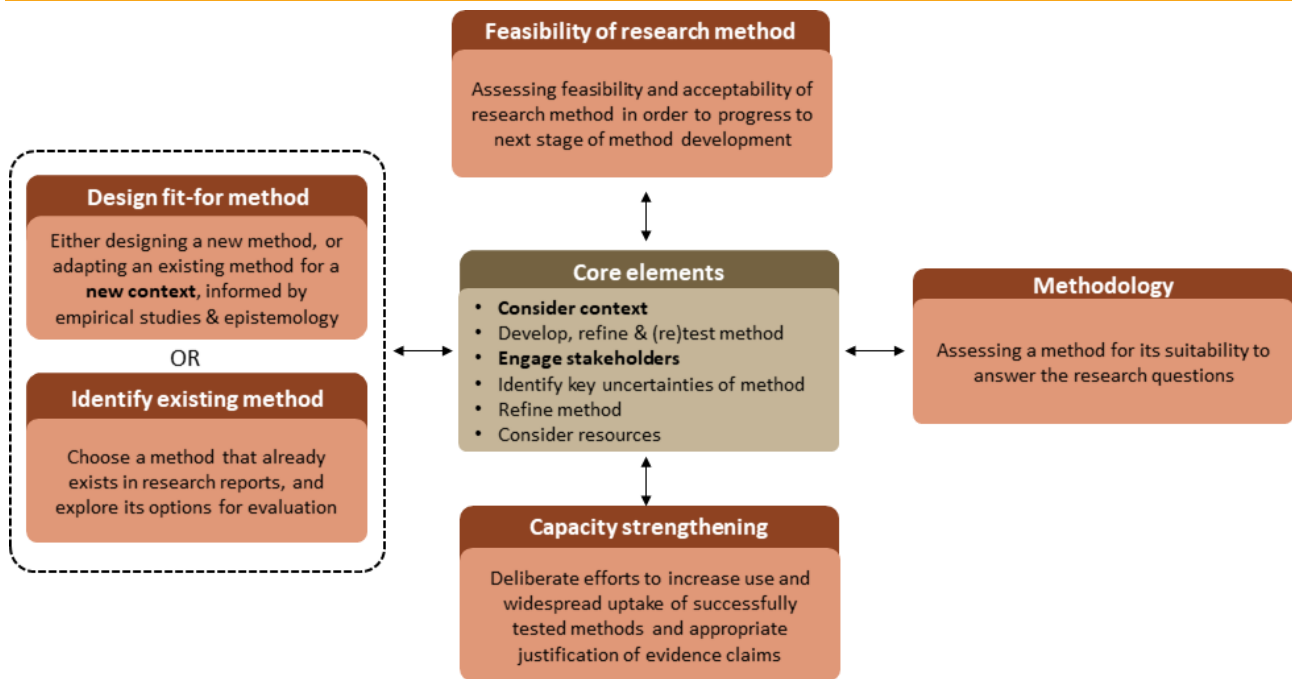


Figure 2: Framework for developing and evaluating research methods

However, these frameworks remain research-led, rather than purposely designed for decision-making. Neither framework considers issues raised by the CEDIL workshop: interventions’ or methods’ relevance, importance or utility, and how they are communicated to decision makers. Nor do they take into account concepts raised by the wider literature (Gough, 2021): the values underlying interventions or evaluation methods.

To address these shortcomings, in our preparatory work (Oliver *et al.*, 2023), we develop a framework (Figure 3) that emphasises the following: the demand for research that strengthens confidence in the evidence claims that matter most for decision-making (as recommended in the CEDIL workshop); and the values and perspectives of decision makers, not only researchers (as recommended by Gough (2021)). The core tasks in the framework begin with articulating evidence needs and engaging stakeholders. The framework balances the conceptualisation of evidence claims with the process of engaging stakeholders. Indeed, the two are necessarily intertwined as stakeholders debate and justify claims, systematically appraising them conceptually and technically, with support from guidance and tools for both technical tasks and the collaborative process. This framework is developed , using a very small set of empirical and methodological studies. Therefore, the questions remain about its wider applicability.

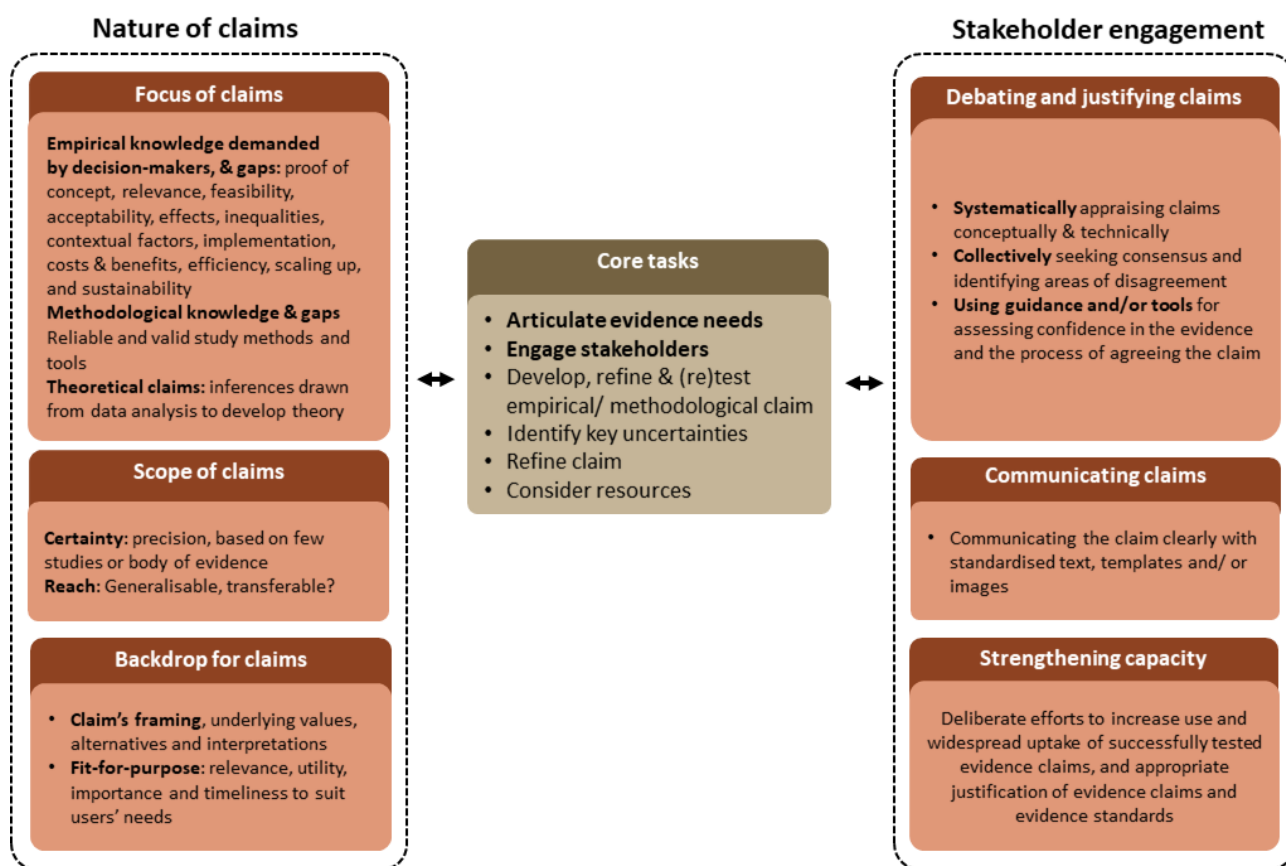


Figure 3: Framework for developing, justifying, and communicating evidence claims

1.5 Research questions

This study aims to explore how evidence claims are supported by impact evaluations and evidence syntheses in the global South, not only in terms of the methodological focus and scope of the claims, but also in terms of the backdrop that frame the claims (illustrated in the left-hand column of the framework in Figure 3). It also explores how evidence claims are debated, justified and communicated in other disciplines, such as medicine, psychology or education (illustrated in the right-hand column of the framework in Figure 3). This exploration is driven by the following questions:

1. What is the nature of evidence claims in impact evaluations and systematic reviews in international development?
2. How are evidence claims developed? What criteria, tools, standards or arguments are used to justify the evidence claims from impact evaluations and evidence syntheses in international development?
3. How are evidence claims communicated for policy and practice?

2 Methods

This chapter presents a summary of the research approaches we employed to address the research questions. This project aims to interrogate recent impact evaluations³ and systematic reviews in international development to understand the nature of evidence claims, how they are made, and to identify evidence standards and tools for producing, supporting and assessing evidence claims.

The research questions are addressed by analysing two sets of studies in parallel (see Figure 4). The first is a set of empirical studies, impact evaluations and systematic reviews of impact interventions that address global South issues. The second is a set of academic texts evaluating and discussing challenges in, and approaches for, making and justifying evidence claims. The learning that comes from analysing the two sets of studies independently is used to identify the following: common practices for justifying evidence claims, their challenges, and arguments for their support; and theoretical proposals for justifying evidence claims, and advances in realising these approaches in practice. The description of practices in regard to making evidence claims revealed by the synthesis of empirical studies (Part A) is supplemented by the narrative description of the literature (Part B).

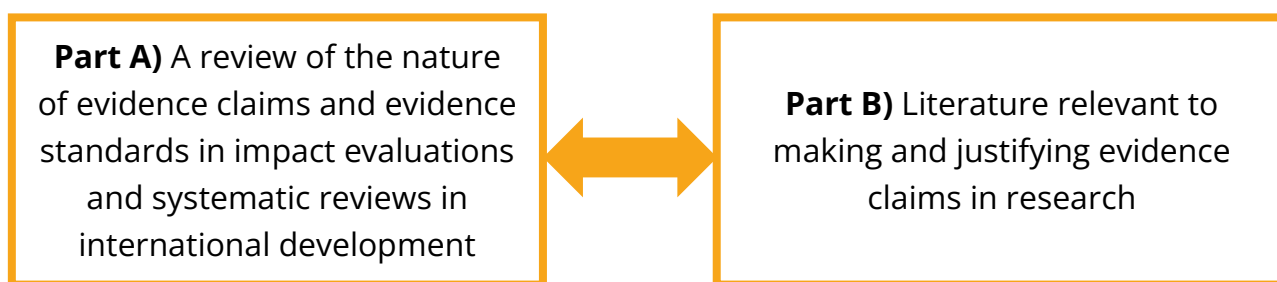


Figure 4: Overall approach of the review

2.1 Identifying studies

We developed explicit criteria to identify relevant literature (see Table 1). For a review of empirical studies (Part A), we employed a purposive sampling strategy, with the aim of identifying a coherent set of empirical studies conducted in different development

sectors. At the initial searching stage, we searched 3ie’s Development Evidence Portal. The EPPI-Centre database was searched. We identified a large set of relevant impact evaluations and systematic reviews. The review protocol was developed as part of the initial searching process described above. After receiving the peer reviewers’ feedback on the protocol, we developed a final set of inclusion/exclusion criteria to narrow down the scope of the review. As a result, we considered impact evaluations and systematic reviews funded by the FCDO that have been published in 2017 or later. Additionally, to be included in the review, studies had to have aimed to assess the impact of intervention programmes in any policy sectors implemented in LMICs using any research designs and synthesis methods. Impact evaluations are defined as *‘an assessment of how the intervention being evaluated affects outcomes, whether these effects are intended or unintended.’*¹⁴ In this review, we considered studies that aimed to assess how the intervention might affect the (un)intended outcomes (Manning *et al.*, 2020). We included only studies published in English (see Table 1). The searches were carried out between August and November 2022, in 3ie’s Development Evidence Portal and the EPPI-Centre website, filtered by year of publication and research funded by FCDO.

	Inclusion criteria	Exclusion criteria
Population	All human populations, at any level of analysis (e.g. individual, family, business, institution, community, settlement, region)	Non-human populations
Interventions	All policy sectors	Studies not assessing interventions
Focus of the study	Studies/reviews that aim to assess the impact of interventions	Studies/reviews that do not aim to assess the effect of interventions (e.g. process evaluation studies or those that aim to evaluate whether an intervention has been implemented as planned or those

¹⁴ <https://www.oecd.org/dac/evaluation/dcdndep/37671602.pdf>

	Inclusion criteria	Exclusion criteria
		that aim to explore barriers to and facilitators of programmes)
Study design/synthesis methods	Empirical studies (e.g. studies with questions, methods and analysis of qualitative and quantitative data)	Not empirical studies (e.g. editorials, commentaries) Literature reviews (no methods, search strategies reported)
Outcomes	Any	None
Context	LMICs	Studies/reviews conducted in high-income countries
Date	In or after 2017	Before 2017
Language	English	Non-English

Table 1: Inclusion and exclusion criteria for impact evaluations and systematic reviews

The inclusion criteria and the coding tools were also further revised to address the reviewers' comments. We piloted the revised inclusion criteria by comparing decisions by two reviewers using an inclusion worksheet with guidance notes. Each report was screened on the basis of titles and abstracts. The review team met to resolve any differences between the reviewers. During the review team meetings, we also acknowledged that the scope and the nature of impact evaluations and systematic reviews addressing global South issues is large and diverse. Therefore, we made the decision to randomly select studies, so as to have a maximum five studies/reviews in each policy sector, with the aim of having a coherent set of literature covering a range of policy sectors within the resources available for the current review. This exercise aimed to strike a balance between gaining an in-depth understanding of the included studies whilst maintaining a wide breadth of identified literature. Full reports were subsequently obtained for the titles and abstracts judged as meeting the inclusion criteria or where there was insufficient information from the title and abstract to assess relevance. In addition, we conducted supplementary searches to ensure that studies focusing on less common policy sectors and/or employing non-RCTs were included in the review. These searches were made in the following:

- Google Scholar;
- Research Development outputs at FCDO (<https://www.gov.uk/research-for-development-outputs>);
- Social Systems Evidence (<https://www.socialsystemsevidence.org/>);
- International Development Research Center (IDRC) (<https://www.idrc.co.uk/>);
- United States Agency for International Development (USAID) (<https://www.usaid.gov/>);
- Norwegian Agency for Development Cooperation (Norad) (<https://www.norad.no/en/front/the-knowledge-bank/>);
- Africa Centre for Evidence (ACE) <https://africacentreforevidence.org/academic-papers/>; and
- Institute of Development Studies (IDS) (<https://www.ids.ac.uk/>).

For Part B, a rapid scoping exercise was undertaken to identify and describe key literature that aims to understand the challenges and implications involved in making and justifying evidence claims. This involved an iterative, selective search of the literature to capture a range of relevant papers, rather than achieving comprehensiveness. Firstly, all key papers which were identified as part of the protocol development were included for screening. These are listed in Appendix D of the protocol (Bangpan *et al.*, 2022) and included those funded by CEDIL, such as CEDIL Lessons Learnt papers, pre-inception and inception papers, seminal papers on evidence-informed decisions, and papers focusing on assessing the quality of research evidence and research use. Secondly, exploratory searches were undertaken during October and November 2022 comprising: a) keyword searches in Web of Science Core Collection and Pubmed; b) related reference searches in PubMed; and c) specific searches of certain papers in Google Scholar using 'cited by'. Thirdly, following screening of the results from exploratory searches further searches were undertaken:

- i) additional papers were identified from the scanning of the papers themselves;
- ii) a 'related-publication search' was undertaken on 44 papers using OpenAlex;
- iii) a citation search was undertaken on Gough (2021); and
- iv) title searches were made in OpenAlex; the OpenAlex searches were undertaken within the EPPI-Reviewer interface and the titles of 524 records identified from steps ii) and iii) were scanned, resulting in identifying 49 items for further screening.

To be included in the current review, the full texts of the papers needed to meet the following criteria: 1) clearly establish, explore, critique or test criteria for making

claims at a meta-level; and 2) be in English, in a format that is easily available to use. The papers were briefly keyworded by one person to note the research field they were published in, their study type, and whether they focused on particular types of claims or research designs, or particular aspects of reporting. To manage the workload, sampling was applied to identify a subset of papers for analysing, privileging: 1) records published within the field of international development; 2) capturing a range of relevant records, rather than aggregating similar records, which was informed by the brief keywording.

2.2 Coding and data extraction

For the review of impact evaluations and systematic reviews, we used the framework for developing, justifying and communicating evidence (Figure 3 in Chapter 1) as a structure for organising and describing the nature of evidence claims in international development. We extracted data to identify the claims and justifications to support the authors' reported claims on the empirical findings, knowledge gaps, methodology and reach. The reviewers extracted data from the introduction, literature review, methods, findings, discussion and recommendation sections using tools developed specifically for this review (see Annex B). We also extracted data, when available, on whether the authors used any other outputs, text or diagram to communicate the relevant claim. The data extraction tools were piloted by all reviewers on a set of the studies in the review to consider if any revisions or additional guidance were needed. The following information was extracted from impact evaluations and systematic reviews:

- **Section A: Bibliographic details:** publication details, date and type of publication, authors' institutions, funders, and any other outputs that highlight good practice or recommendation.
- **Section B: Population characteristics:** age, gender, other characteristics (such as those with health conditions or focusing on disadvantaged populations), as specified by the study.
- **Section C: Study characteristics:** study aims and research questions, theories underpinning the research, geographical location, development policy sectors.
- **Section D: Intervention characteristics:** type of interventions; theory of change or logic model used.
- **Section E: Research methodology:** study design, data analysis and synthesis approaches.

- **Section F: Research gaps:** considering any discussion of previous research, theoretical standpoints and reasonings of evidence producers, equity considerations and evidence users' perspectives, guideline or tools.
- **Section G: Evidence claim:** research methodology – considering the appropriateness and technical quality of the research; evidence standards applied to making the evidence claims (e.g. appropriateness of research methods, robustness of the research, theoretical standpoints and reasonings of evidence producers, ethics, equity); any guideline or tool applied to justify the appropriateness (choice) or technical quality (robustness) of the research; approaches used to communicate the claim.
- **Section H: Evidence claim:** empirical findings – the main empirical findings reported in the abstract, with the relevant information on the certainty of the claim, considering the appropriateness and technical quality of the research to justify the claim (e.g. validity of inferences, precisions of statistic measures, credibility, representation of reality); strengths and limitations of the research that might be undermining or strengthening the claim; other evidence standards applied to make or justify claims from empirical findings.
- **Section I: Evidence claim:** reach – generalisability and transferability.
- **Section J: Evidence claim:** framing – underlying values, interpretations and alternatives. This covered how the authors reported the framing of their work (e.g. drawing on existing literature and/or engaging stakeholders in discussion) and the framing that was apparent in their work (e.g. the values shaping the analysis or having a high profile in the findings).

We piloted the data extraction tools. Two review team members each extracted data from the included studies independently. For Part B, a rapid scoping exercise, all sampled literature were coded on the following: the target audience of the paper; the purpose of the paper; how the justification of claims was analysed; any focus on an aspect of justifying evidence claims; and if there were considerations of values underpinning the claims. This coding provided the basis of the summary tables.

2.3 Analysis approaches

Data extracted from the included impact evaluations and systematic reviews were synthesised using a framework synthesis, where the initial framework was constructed to take into account the key concerns of the main stakeholders (in this case the producers of evidence) and refined as the reviewers became familiar with key concepts arising from the studies being reviewed (Brunton *et al.*, 2020, Oliver *et al.*, 2005). The reviewers read and re-read data and applied line by line codes to capture and interpret the meaning and organise the data and coding (e.g. the nature of

evidence claims, approaches used to reach conclusions, evidence standards) into the three main types of evidence claim domains: i) choice of research methodology; ii) empirical findings, iii) reach. We narratively described the key characteristics of impact evaluations and systematic reviews included in the review (also see the tables in Annex D). Our analysis was based on the data presented in the detailed summary tables of the included impact evaluations and systematic reviews (see the supplementary file).

The analysis also explored the key characteristics of research by policy sectors, types of research questions (substantive or methods questions), or types of data analysis (e.g. statistical analysis, meta-analysis, narrative synthesis) to provide a more fine-grained understanding of the different types of evidence claims and how they were reached and justified. For the theoretical papers, we described key elements relevant to evidence claims and evidence standards reported, and mapped these into overarching papers, and those that cover one or more of the three types of evidence claim domains, and additionally the domain of communication.

Eight potential values were identified in advance from a review of literature aiming to identify and organise the variety of different viewpoints in the literature on social values in health and care guidance (Gough *et al.*, 2014). Other values were recorded as they were recognised in the studies. The analysis of values was the most innovative aspect of this work and was applied to the systematic reviews by a single author, as a proof of concept analysis. The findings of this analysis are presented in Annex E.

Finally, we investigated the relationship between the nature of evidence claims and how they had been developed by constructing a 'truth table' that cross-tabulated the different elements of our evidence claims framework (Annex G). For each study, we recorded and analysed the relationships between the following four sets of attributes:

- stakeholder engagement for debating the focus, scope of backdrop of the review (either for individual reviews as part of earlier consensus development exercises);
- social values integrated in the review (as components of the intervention or theory of change, or outcomes of interest, or elsewhere in the text);
- the certainty or reach of evidence claims (in terms of the quality of the underpinning studies and the transferability of the findings); and
- the social impact of reviews – in other words, where they appeared and influenced the world beyond research.

3 Making and justifying claims in international development studies

To address the review questions, we included 78 studies in the review. We narratively presented key themes based on the data summarised (see the data extraction tables in the supplementary file). The organisation of this chapter is guided by the framework for developing, justifying and communicating evidence (Figure 3). First, we describe the focus of the empirical and methodological claims in impact evaluations and systematic reviews (Section 3.1). Next, we present the scope of the claims in terms of how the findings are generalisable or transferable (Section 3.2). We also discuss how the authors communicate the claims (Section 3.3) and how evidence claims were framed (Section 3.4).

3.1 Focus of the claims

For this review, we explored the focus of the claims in the impact evaluations and systematic reviews in terms of the empirical knowledge claims and the methodological claims. Within these two literatures, we also explored the backdrop to claims, in terms of the values expressed by the study authors, in order to understand how evidence claims were framed and communicated to fulfil their purpose of informing decisions. The sections below are structured as follows: Section 3.1.1 discusses empirical knowledge claims of impact evaluations; Section 3.1.2 discusses empirical knowledge claims of systematic reviews; Section 3.1.3 discusses methodological claims of impact evaluations; and Section 3.1.4 discusses methodological claims of systematic reviews.

3.1.1 Empirical knowledge claims of impact evaluations

Empirical knowledge claims were extracted from the abstract or main summary of all studies. Claims about intervention impact were the most common, being present in all 47 studies. A further 23 studies also made a claim about the magnitude of this impact: for example, through reporting the effect size. To analyse the justification for these claims, we extracted data where authors considered the certainty and strength of their findings based on the technical quality of their research, or the consistency and appropriateness of the research. Strengths and limitations which might be undermining or strengthening empirical knowledge claims were also discussed. Empirical claims and their justifications are presented in more details in Table 2 in the supplementary file.

Making and justifying claims about intervention impact

All 47 studies which contained a quantitative element made claims about interventions impacts. The most common justification for the intervention impact claims was the use of standard measures and tests for statistical significance (n=47). All studies adopted academic conventions to prove **the certainty and strength of their claims**, using statistics such as p-values, F tests and confidence intervals. The reporting of significance within text ranged from shorter summaries, such as discussing a '*statistically significant effect*' (Ahmed *et al.*, 2019, p. 14) to explicit reporting of these values, such as '*This includes: the number of active members per household (T = 4.25, C = 4.39; p= 0.382), the age of head of household (T = 49.0, C = 47.5; p = 0.245) or even the household income (p = 0.768)*' (Béné *et al.*, 2020, p. 7)

Justifications were often rooted in methodological quality (n=31). The most common of these justifications was the use of robustness checks to validate the certainty of claims. A direct example of this is Hirvonen and colleagues, who '*assessed the robustness of this finding in several ways*' (Hirvonen *et al.*, 2017 p. 310). One study made a direct link between robustness and the trustworthiness of evidence claims: '*to be certain our results are trustworthy, we need to account for this threat of spillover effect*' (Parker *et al.*, 2019, p. 47). Of the 15 studies which used the robustness of results as a justification for empirical claims, 13 achieved this through the use of explicit statistical modelling techniques, such as sensitivity analysis, testing additional control variables and robust standard errors. Two studies, Béné *et al.* (2020) and Cocciolo *et al.* (2020), presented their results as robust, based on the general rigour and quality of their methods and data collection. Similarly, eight studies checked potential biases or performed balance checks to prove the effective randomisation and internal validity of the evaluation (Aker *et al.*, 2017; Angeles *et al.*, 2019; Armand *et al.*, 2019; Bandiera *et al.*, 2017; Barnett *et al.*, 2018; Diagne and Cabral, 2017; Källander *et al.*, 2021; Pellegrini, 2018). Several studies also conducted sub-group analysis to assess heterogeneous effects (Gibbs *et al.*, 2020; Khwaja *et al.*, 2020; Najy *et al.*, 2018; Roth *et al.*, 2017; Saboya *et al.*, 2018; Zegarra *et al.*, 2017).

Another common way authors justified their claims about empirical findings was through exploring their **consistency** with published literature (n=22). Where similar, findings were defined as aligned with or '*consistent with previous studies*' (Rakotonarivo *et al.*, 2017, p. 487). The magnitude of findings was often contextualised through this comparison with other studies: for example Belissa and others (2019, p. 277) explicitly compared effect sizes between studies, noting that '*The effects sizes we find are smaller than those of Casaburi and Willis (2018)*'. Inconsistency with existing research was also

highlighted where relevant: for example, *'in contrast to research that suggests that some women's economic empowerment programming can increase women's experiences of IPV, there was no suggestion that this occurred'* (Gibbs *et al.*, 2020, p. 14).

Claims were also justified through their consistency with authors' expectations, predictions and intervention hypotheses (Asfaw *et al.*, 2017; Brudevold-Newman *et al.*, 2017). Furthermore, theory was used to justify the certainty and impact of claims across several studies (n=7). This was done through a return to the theory of change (Béné *et al.*, 2020; Cocciolo *et al.*, 2020; Parker *et al.*, 2019) or an examination of the consistency of results with established theories (Amirapu *et al.*, 2020; Belissa *et al.*, 2019; Berhane *et al.*, 2017; Gibbs *et al.*, 2020).

We found attempts to justify the certainty of claims through the exploration of reasons for, or alternative explanations behind, results within 19 studies. For example, Ejekumhene and others (2019) discuss potential confounders and history effects which *'refer to events that happen in the environment that change the conditions of a study, affecting its outcome'* (p. 64). Morten and colleagues (2020) provided guidance to readers on interpreting the results, based on potential explanations, providing *'three caveats to these findings'* (p. 24). These attempts to provide alternative explanations or draw on theory were particularly common across eight studies that suggested a lack of intervention impact due to insignificant results (Banerjee *et al.*, 2018; Barnett *et al.*, 2018; Berhane *et al.*, 2017; Gibbs *et al.*, 2020; Morten *et al.*, 2020; Mvukiyehe and van der Windt, 2020; Pellegrini, 2018; Tsai *et al.*, 2018).

For mixed-methods studies with a qualitative element, triangulation was used to provide further evidence for empirical claims (n=8). For example: *'these multi-layered triangulation processes strengthen confidence in the findings'* (Barnett *et al.*, 2018, p. 44); *'Overall, there is strong quantitative and qualitative evidence that participation in the WfWI economic and social empowerment intervention incrementally strengthened women's livelihoods and economic well-being'* (Gibbs *et al.*, 2020, p. 14); and *'This is consistent with qualitative interviews conducted with informal businesses not targeted by the program'* (Najy *et al.*, 2018). One study, Avdeenko and Frölich (2019) presented direct quotations from participants alongside its main results section .

Of the 23 studies which made an intervention impact claim, presenting the magnitude of the impact within the abstract or summary, 20 justified claims with examples of the size of the intervention's effect, such as *'the standard adult education program increased their math and reading test scores by 0.19–0.22 s.d, respectively'* (Aker and Ksoll, 2019, p. 239). A further 14 studies provided an explanation of the meaning of the effect size, for example explaining that *'the effect size is tiny at around 0.3 per cent'*

(Angeles *et al.*, 2019, p. 27), or the effect is *'relatively small'* (Banerjee *et al.*, 2018, p. 127), or *'economically large'* (Brudevold-Newman *et al.*, 2017, p. 1). Piper and others (2018) provided explicit guidance for the interpretation of effect sizes, drawing on published guidelines to determine appropriate benchmarks: *'utilizing Cohen's guidance, the DFID PRIMR treatment groups had significant effects on learning outcomes'* (Piper *et al.*, 2018) (p. 330).

One study also used ethical research standards to contextualise the results and provide an explanation for **the appropriateness of its findings**: *'But to be safe (and avoid poisoning due to over-consumption of iron), the concentration of iron supplementation in the food must be limited...This of course further reduces impact, and ultimately makes the strategy non-viable'* (Banerjee *et al.*, 2018, p. 130).

We identified an explicit reference to any **strengths and limitations** which might be undermining or strengthening empirical claims in 37 studies. 20 studies highlighted the strengths of their research, with 31 highlighting limitations. Strengths were exclusively related to methodological quality, with the most common relating to an acknowledgement of biases within the study design (n=11) and the robustness of the study (n=7). Armand and others (2019) discussed how they *'supplement survey-based measurements with behavioural measurements to reduce the risk that survey-based measurements alone could be biased by Hawthorne effects'* (p. 13). Belissa's (2019) confirmation of successful balance checks, *'Tables 1 and 2 suggest the randomization worked well'* (p. 272), is an example of studies highlighting effective randomisation to strengthen their empirical claims. Similarly, Mvukiyehe and van der Windt (2020) used the study sample size to stress the validity of claims: *'Second, the size of the experiment we study here is much larger...significantly increasing statistical precision and decreasing the possibility for Type-II errors'* (p. 14). Five studies used the general strengths of their chosen study designs or intervention techniques as evidence of strong research (Avdeenko and Frölich, 2019; Bandiera *et al.*, 2017; Morten *et al.*, 2020; Parker *et al.*, 2019; Robertson, 2019).

The limitations that were cited were also largely methodological, with the majority (n=20) relating to biases or baseline imbalances. Many included references to the impact of social desirability bias: for example, Cocciolo and others (2020) acknowledged *'previous research (e.g. Ahuja *et al.* 2010) and our own experience suggest that social desirability bias influences household reports of behaviour with respect to obtaining drinking water'* (p. 36). An inability to make certain causal claims was mentioned as a limitation in four studies (Carew *et al.*, 2020; Chirwa *et al.*, 2017; Morten *et al.*, 2020; Piper *et al.*, 2018). Limited external validity or small sample size was reported in five studies (Aker and Ksoll, 2019; Coleman *et al.*, 2019; Harris-Fry *et*

al., 2018; Källander *et al.*, 2021; Saboya *et al.*, 2018). Several studies referenced potential confounders or difficulty isolating individual effects of the intervention, limiting the empirical claims made about its impact or magnitude. For example, Béné *et al.* (2020) discussed the potential confounding effects of existing interventions in the same area, while Gibbs and others (2020) acknowledged ‘*we do not know if it was the vocational training, cash transfer given during training or both that benefited women economically. Our hypothesis is that both were important but we have not tested this*’ (p. 16).

3.1.2 Methodological claims of impact evaluations

We identified methodological claims in 34 impact evaluation studies. The majority of these evaluations made claims about the choice of study design (n =13), specific methods for data collection, and data analysis (n=18). Claims about the quality and validity of the study and claims about innovative methods were made in six studies each. Three studies made claims about the usefulness of the chosen research design. Methodological claims and their justifications are presented in more detail in Table 1 in the supplementary file.

Making claims about the choice of study design

Of 13 studies making claims about the choice of the chosen study design, seven employed quantitative research methods, whereas six studies employed a mixed-methods design. The choice of the study design was often supported by the **expected quality or added value** of employing the chosen research design to address the research questions on causality (Avdeenko and Frölich, 2019; Bandiera *et al.*, 2017; Brudevold-Newman *et al.*, 2017, Cocciolo *et al.*, 2020). For example, one mixed-methods study evaluating community building and humanitarian aid interventions in Pakistan explained the benefits of the RCT for establishing causality based on the post-intervention differences. At the same time, the authors of three studies claimed that qualitative research had an explanatory power to corroborate the observed quantitative findings, clarify possible contradictions in quantitative results, contextualise the quantitative findings in their cultural and local circumstances, and gain a deeper understanding of participants' perceptions (Avdeenko and Frölich, 2019; Berhane *et al.*, 2017; Parker *et al.*, 2019).

Approaches to address the issues of the counterfactual were one of the main justifications when making decisions on the study design. Implementation constraints such as the time and resources available, local conditions, ethical concerns, sample size, costs, and data availability, were often cited as factors when considering whether

it was **feasible and appropriate** to conduct RCTs or other types of quasi-experimental methods (Angeles *et al.*, 2019; Barnett *et al.*, 2018; Béné *et al.*, 2020; Berhane *et al.*, 2017; Pellegrini, 2018; Roth *et al.*, 2017; Saboya *et al.*, 2018; Zegarra *et al.*, 2017). Béné and colleagues (2020) discussed the objectives, strengths, weaknesses, and applications of different quasi-experimental designs in relation to their research. They specified that:

'no single design is considered "best" instead the choice usually depends on a number of factors including the purpose of the evaluation, how long the evaluation has been planned, if data from before the programme started (baseline) is available, whether there are data or resources for data collection at several measuring points along the way in the programme or whether there may be a possibility to generate a control or a relevant comparison group for the intervention group [6,7].'

Similarly, Angeles and others (2019) evaluated a social cash transfer programme in Zimbabwe and justified their study design choice by explicitly explaining the reasons for conducting a *'quasi-experimental design rather than a randomised controlled trial'*, relating to the challenges of local policy, ethics, and the feasibility of constructing a comparable control group;

'A major factor in the choice of a quasi-experimental design rather than a randomised controlled trial is the stated policy of the Ministry that all eligible households will be enrolled in the programme once a district enters the programme. In other words, the programme will immediately be scaled up to full coverage within each district. The Ministry determined that it would be ethically and politically unfeasible to provide the programme to some households while delaying others within the same district to serve as a control group because it would conflict with this stated policy. Therefore, a randomised controlled trial design was not possible because all eligible households within a district must receive the programme at the same time.' (Angeles *et al.*, 2019)

Making claims about the choice of research methods

We identified claims about the choice of research methods in 18 impact evaluations. The decisions made about specific research methods were relevant to sampling strategies (e.g. sample size, stratification), data collection (e.g. choices of variables or measures, the duration of data collection, sources of data), and data analysis methods

(e.g. intention-to-treat analysis, choices of modelling). The choice of the chosen research methods was justified by the appropriateness in regard to minimising biases (Armand *et al.*, 2019; Asunka *et al.*, 2019; Blattman *et al.*, 2018; Chirwa *et al.*, 2017; Piper *et al.*, 2018) and addressing implementation challenges (Aker and Ksoll, 2019; Banerjee *et al.*, 2018; Hirvonen *et al.*, 2017; Källander *et al.*, 2021; Pellegrini, 2018). Other justifications included acknowledgement of the fact that these methods were commonly used in the field, or based on previous work (Asfaw *et al.*, 2017; Bett *et al.*, 2018; Gibbs *et al.*, 2020; Tsai *et al.*, 2018). Three other studies argued that the chosen methods were designed to encourage participation (Cocciolo *et al.*, 2020), be adaptable (Li and Liu, 2020), and fill the research gaps (Morten *et al.*, 2020). Four studies claimed that the chosen research methods allowed them to address spillover effects. The authors explained that spillover effects can pose a challenge when interpreting results as they can take multiple forms with different sizes and directions, increasing the risk of bias of the treatment effects (Asunka *et al.*, 2019; Blattman *et al.*, 2018; Coleman *et al.*, 2019; Parker *et al.*, 2019).

Making claims about quality and validity

Claims about the quality and validity of the chosen methods were made in 14 studies. Several studies implied that the nature of randomisation could make studies rigorous and relevant (Asfaw *et al.*, 2017; Blattman *et al.*, 2018; Källander *et al.*, 2021; McKenzie, 2017; Parker *et al.*, 2019; Pellegrini, 2018; Piper *et al.*, 2018), particularly when the availability of evidence is limited, such as in a conflict-affected setting (Gibbs *et al.*, 2020). To justify claims about the quality and validity of the study findings, the authors ensured that their data collection and data analysis were appropriate, sensitive to local contexts (Angeles, 2019), addressed ethical issues, mitigated potential biases (Avdeenko and Frölich, 2019; Pellegrini, 2018), avoided overstating precision (Blattman *et al.*, 2018), and triangulated data from different sources or methods (Barnett *et al.*, 2018; Dar *et al.*, 2020; Saboya *et al.*, 2018). One study justified a claim about the study quality by stating that they employed validated research instruments and had an appropriate sample size for the data analysis (Hirvonen *et al.*, 2017).

The choice of the chosen research methods was justified by the fact that these methods offered alternative approaches for data collection and data analysis, in six studies (Aker *et al.*, 2017; Chirwa *et al.*, 2017; Diagne and Cabral, 2017; Hirvonen *et al.*, 2017; Li and Liu, 2020; Robertson, 2019). For example, the authors explained that they '*employed multiple imputation by chained equations as an alternative robustness exercise for panel attrition*' (Aker *et al.*, 2017, p. 199) or '*used alternative measures of nutrition*

knowledge' (Hirvonen *et al.*, 2017), or that *'the difference-in-difference approach provides a valid alternative to the PSM approach...'* (Robertson *et al.*, 2019).

Fewer studies made claims about the innovative nature of their chosen methods (Armand *et al.*, 2019; Asfaw *et al.*, 2017; Béné *et al.*, 2020; Cocciolo *et al.*, 2020; Hirvonen *et al.*, 2017) and their usefulness to inform policy and contribute to research (Armand *et al.*, 2019; Barnett *et al.*, 2018; Blattman *et al.*, 2018). The research methods were considered innovative when they were not commonly used in the field or not easy to apply or implement. One study claimed that using a novel data set in the study allowed it to assess market access, which explained the reasons for variations in the observed findings regarding nutrition knowledge and children's diet in Ethiopia (Hirvonen *et al.*, 2017). Béné and others (2020) explained that their chosen statistical evaluation approach was *'one step beyond most other resilience analyses currently found in the literature'*. Another study stated:

'Our procedures for determining these matches are novel, because the problem of linking households to decentralised infrastructure is not easy to solve' (Cocciolo *et al.*, 2020)

3.1.3 Empirical knowledge claims of systematic reviews

We also explored the empirical knowledge claims made by the reviews' authors in the abstract or executive summary sections, which were supplemented by claims and justifications discussed in the sections on methods, limitations and strengths, and conclusions in the main reports (when available). The majority of the systematic reviews made claims about intervention impact (n=27). Further, 17 reviews made claims about research gaps, while 14 made claims about implementation success or failure. One review made a claim relevant to mechanisms that linked contextual factors and outcomes. Of 31 reviews, 23 justified the nature of the body of evidence in terms of the quality, quantity, consistency and focus when reporting empirical findings.

Making claims about intervention impact

When claims were made about intervention impact, the review authors often described the degree of certainty of the body of evidence they considered in the review. Their justifications often included **the magnitude** of the impact, by reporting the effect sizes of individual primary studies or pooled effect sizes (Alampay *et al.*, 2017; Ilavarasan *et al.*, 2017b; Langer *et al.*, 2018; Waddington *et al.*, 2019; Wolf *et al.*, 2018). Other justifications for claims about intervention impact were **the size of the**

evidence base (Catalano Richard *et al.*, 2019; Maynard *et al.*, 2017; Nair *et al.*, 2017a; Nidhi *et al.*, 2017; Stone *et al.*, 2020), **the availability of the evidence** (Kumar *et al.*, 2018), **the consistency** of the evidence across the effectiveness outcomes (De Buck *et al.*, 2017; Pilkington *et al.*, 2017; Waddington *et al.*, 2019), **the focus of the evidence base** (such as population focus, geographical locations or characteristics of interventions) (Hossain *et al.*, 2017a; Nair *et al.*, 2017a), and **the overall quality of a body of evidence** (Catalano Richard *et al.*, 2019; Langer *et al.*, 2018; Nair *et al.*, 2017a; Obuku *et al.*, 2017; Yount *et al.*, 2017). Seven reviews reported inconclusive findings, with the authors describing the body of evidence as *'mixed, with few significant outcomes'* (Hossain, 2017a, p. 2). Others justified the inconclusive findings because of the strength of evidence, which was limited (Akparibo *et al.*, 2017), and the inconsistency of the findings across the included studies (Williamson *et al.*, 2017). Four studies reported non-statistically significant findings, where the body of evidence was described as insufficient or limited (Ilavarasan *et al.*, 2017b; Kumar *et al.*, 2018; Stone *et al.*, 2020).

Making claims about research gaps

We identified 17 reviews that made explicit claims about knowledge gaps. The justification for these claims was not always explicitly stated. Four reviews made claims about knowledge gaps that could be related to **inconclusive evidence or inconsistent findings** from reviews (Akparibo *et al.*, 2017; Alampay *et al.*, 2017; Hossain *et al.*, 2017a; Menon *et al.*, 2018). Several reviews acknowledged **a paucity of research** conducted with specific marginalised groups, in conflict settings, or relating to the cost effectiveness of the intervention (Catalano Richard *et al.*, 2019; De Buck *et al.*, 2017; Langer *et al.*, 2018; Nair *et al.*, 2017b; Obuku *et al.*, 2017; Wolf *et al.*, 2018; Patel *et al.*, 2017). For example, De Buck and colleagues (2017) conducted a mixed-methods review, stating that *'no studies were performed in a disaster setting, and more research specific context is warranted'*.

When knowledge gaps were identified, they were often used to justify future research recommendations. The authors who made claims about insufficient evidence or inconsistent findings frequently recommended conducting future research with more in-depth and exploration of the context for which evidence was lacking (Akparibo *et al.*, 2017; Blundo-Canto *et al.*, 2018; Hossain *et al.*, 2017b; Ilavarasan *et al.*, 2017a; Menon *et al.*, 2018; Nair *et al.*, 2017b; Patel *et al.*, 2017; Pilkington *et al.*, 2017). When a claim about knowledge gaps related to the inadequate quality of the body of evidence, future rigorous research was frequently recommended (Nair *et al.*, 2017b; Nidhi *et al.*, 2017; Obuku *et al.*, 2017; Patel *et al.*, 2017).

Making claims about programme implementation

We identified 12 reviews that made claims about the factors influencing programme implementation. Nearly all reviews (n=11) included different types of study design, including qualitative and mixed-methods research. The review authors acknowledged that qualitative investigation would allow further exploration, by collecting rich data to understand how interventions were implemented. However, the specific justifications for claim about implementation factors were generally unclear. There was an implicit assumption that analysing qualitative data as part of the review process could elicit insights into the implementation process to inform decisions on programme design and delivery. Information on **the availability, and the size, of** studies investigating factors influencing implementation was reported in four reviews (De Buck *et al.*, 2017; Kumar *et al.*, 2018; Maynard *et al.*, 2017; Nidhi *et al.*, 2017). Poor reporting quality, including limited information on programme descriptions and a lack of qualitative research, were referred to as limitations in making evidence claims about implementation factors in two reviews (Akparibo *et al.*, 2017; Catalano Richard *et al.*, 2019). One review, which included only experimental studies, was able to identify key elements of a successful intervention by examining studies with '*significant results in consideration to the context where these [the key elements] were applied*' (Hossain *et al.*, 2017a). One evidence synthesis used a fishbone diagram as a synthesis approach to identify key programme elements that had 'strong evidence' linking them to effective interventions (Annamalai *et al.*, 2017).

Justifying empirical knowledge claims: technical quality, perspectives, and strengths and limitations of systematic reviews in international development

To explore how evidence claims were made, we considered how research findings were brought together and justified (Gough, 2021). First, we outline the justifications made by reviewers relating to the technical quality of the execution of systematic review methods. This involved considering whether the review authors used specific tools to guide a review process and appraise the quality of the included studies, and if they applied a clear approach to assess the certainty of the evidence base. Of 31 reviews, 23 developed a review protocol to define the review scope and to outline review approaches in advance, and 13 reviews worked with external research teams during the key stages of the review process. All but one review assessed the quality of the included studies. Four reviews used Grading of Recommendations Assessment, Development and Evaluation (GRADE) to assess the certainty of the evidence base (De Buck *et al.*, 2017; Garn *et al.*, 2017; Langer *et al.*, 2018; Obuku *et al.*, 2017).

Secondly, we considered the values and perspectives of the review authors, experts outside the research team, and policymakers, as these might have shaped the review process and how evidence claims were made. The authors of three reviews acknowledged the importance of knowledge generated from qualitative research. They argued that qualitative data provided rich and insightful accounts to inform programme design. They adapted synthesis methods by bringing together the 'best available evidence' in the absence of a quantitative research study (De Buck *et al.*, 2017; Ghose *et al.*, 2017; Langer *et al.*, 2018). At the same time, the authors of two reviews valued the team's perspectives in the absence of high-quality evidence in the field (Patel *et al.*, 2017; Zwi *et al.*, 2018).

'A major finding of this systematic review is the striking lack of high quality evidence on targeting vulnerable populations in urban humanitarian emergencies. The research team's own experience of assessing the strength of the evidence in many papers revealed...'. (Patel, 2017)

The authors of one review revealed different perspectives on what constituted 'valuable evidence' and how different methodological approaches could be applied. Williamson and others (2017) were concerned that some excluded papers were 'relevant' and had 'interesting' observations but did not meet the review's eligibility criteria, meaning a wealth of knowledge was missed that could inform policy and practice in the field.

Perspectives from experts outside the review teams were considered in 20 reviews. These experts were involved in the review process in different ways. They provided guidance and feedback to the review teams at various stages of the review by serving as advisory group members (Ali *et al.*, 2017; Annamalai *et al.*, 2017; Ghose *et al.*, 2017; Menon *et al.*, 2018; Nair *et al.*, 2017a; Nair *et al.*, 2017b; Obuku *et al.*, 2017; Pilkington *et al.*, 2017; Zwi *et al.*, 2018). Several reviews used expert inputs to frame the review scope (Catalano Richard *et al.*, 2019; Langer *et al.*, 2018; Maynard *et al.*, 2017), identify relevant literature (Akparibo *et al.*, 2017; De Buck *et al.*, 2017; Nidhi *et al.*, 2017), develop the search strategy (Babu *et al.*, 2017), draw out policy and practice implications (De Buck *et al.*, 2017; Stone *et al.*, 2020) and disseminate the review findings (Kumar *et al.*, 2018).

The authors of seven reviews considered policy perspectives and priorities in order to produce policy-relevant reviews (Annamalai *et al.*, 2017; Babu *et al.*, 2017; Ilavarasan *et al.*, 2017b; Kumar *et al.*, 2018; Langer *et al.*, 2018; Menon *et al.*, 2018; Obuku *et al.*, 2017; Stone *et al.*, 2020). As the authors of one review explained, 'based on the draft

report findings we will be interacting with policymakers to elicit their views on the findings, which, if necessary, will be included in the final report' (Babu *et al.*, 2017).

Finally, we explored the strengths and limitations, discussed in the reviews, which could undermine or strengthen empirical claims. Of 31 reviews, 15 highlighted the strengths of the reviews, with 26 highlighting the limitations. A quality assessment exercise conducted to determine the quality, risk of bias, validity and reliability of the included studies was commonly cited as a strength by the review authors (Garn *et al.*, 2017; Hossain *et al.*, 2017a; Hossain *et al.*, 2017b; Langer *et al.*, 2018; Obuku *et al.*, 2017; Stone *et al.*, 2020). The review authors also considered the importance of the review in filling research gaps in under-researched areas as one of their strengths (Garn *et al.*, 2017; Hossain *et al.*, 2017a; Maynard *et al.*, 2017; Stone *et al.*, 2020; Waddington *et al.*, 2019; Zwi *et al.*, 2018). Five review authors pointed out the benefits of conducting reviews with a broad focus, which allowed the review to identify relevant studies, covering a wide range of interventions, contexts and outcomes, often over a long period of time (Babu *et al.*, 2017; Kumar *et al.*, 2018; Obuku *et al.*, 2017; Stone *et al.*, 2020; Yount *et al.*, 2017). Another four studies emphasised the importance of the quality assurance process, such as performing double screening and coding exercises, or having a quality assurance team to provide feedback and guidance (Garn *et al.*, 2017; Hossain *et al.*, 2017a; Hossain *et al.*, 2017b; Langer *et al.*, 2018). Other strengths of reviews considered by review authors were the fact that the review was comprehensive and employed a robust approach (De Buck *et al.*, 2017; Hossain *et al.*, 2017a; Hossain *et al.*, 2017b; Obuku *et al.*, 2017; Zwi *et al.*, 2018), engaged with key stakeholders in the review process (Hossain *et al.*, 2017a; Hossain *et al.*, 2017b; Langer *et al.*, 2018), considered both quantitative and qualitative data (Stone *et al.*, 2020; Yount *et al.*, 2017), or had a high to medium quality of evidence to answer the review questions (Hossain *et al.*, 2017a; Waddington *et al.*, 2019).

Five main limitations the review authors commonly cited were the heterogeneity of the evidence base in terms of the context, type of interventions, study design or outcome variable (n=18); the poor methodological quality (n=10) and the poor reporting quality (n=11) of the included studies; a lack of the appropriate type of study design or data to address the review questions (n=13); and the potential for publication bias (n=11). Other limitations discussed in the reviews included inconsistent results (n=2), a small evidence base (n=6), and a broader review scope (n=3).

3.1.4 Methodological claims of systematic reviews

We identified methodological claims in 28 systematic reviews. The majority of the studies made claims about the choice of review methods, including the development

of the search strategy, the choice of data quality assessment tools, and synthesis (n=24). The choice of the review approach was discussed in seven reviews. Two reviews made claims related to quality and validity. One review (Waddington *et al.*, 2019) claimed that the review applied an innovative review method to address the review questions. Further information on methodological claims and their justification are presented in Table 4 in the supplementary file.

Making claims about the choice of review approach

While the seven studies making claims about the choice of the review approach had unique designs, there were some similarities in the way the decision was justified. Three of the studies were directed by the heterogeneous nature of the evidence base, which informed their decision to adopt more flexible approaches to accommodate the studies they were synthesising (Langer *et al.*, 2018; Patel *et al.*, 2017; Zwi *et al.*, 2018). Two other study design choices were motivated by the policy needs. Annamalai *et al.* (2017) and Langer *et al.* (2018) developed an evidence map in order to respond to the expectations of evidence users in the field and the emphasised importance of evidence-based policy. Four reviews chose the review approach in order to conform to a conventional standardised procedure for conducting systematic reviews (Langer *et al.*, 2018; Nidhi *et al.*, 2017; Obuku *et al.*, 2017; Yount *et al.*, 2017). Another two review designs were justified by their appropriateness in relation to the research questions. For example, Obuku *et al.* (2017) conducted a systematic review with an aggregative design to align with a research question assessing the effectiveness of an intervention.

Making claims about the choice of research methods

Claims about the choice of the research methods were the most common type of methodological claim in a set of systematic reviews we identified. The types of decisions made relating to the claim included the choice of search strategy methods, the inclusion criteria, the choice of outcomes measured, the choice of quality assessment tool, and the synthesis approach. The claims made about choice of inclusion criteria and the specific search strategy were mostly implicitly justified using arguments about the reviewers' knowledge of the research area and personal expertise (Catalano Richard *et al.*, 2019; De Buck *et al.*, 2017; Obuku *et al.*, 2017), or explicitly justified by reference to external expert opinion (Ilavarasan *et al.*, 2017a). Two claims about the need for triangulation were justified by the nature and type of studies included in the reviews (Nair *et al.*, 2017a; Stone *et al.*, 2020). As one of the review authors stated:

'The triangulation of findings from different research methods allowed us to define and test hypotheses using different methodologies that informed and supplemented each other.' (Stone *et al.*, 2020)

The choice of synthesis approach varied between more aggregative methods, such as meta-analysis (Akparibo *et al.*, 2017; Alampay *et al.*, 2017), and more configurative methods, such as narrative synthesis (Maynard *et al.*, 2017; Williamson *et al.*, 2017), framework analysis (Hossain *et al.*, 2017b, Nair *et al.*, 2017b), textual narration (Kumar *et al.*, 2018), or a combination of the two (Babu *et al.*, 2017). Most often, the authors explained that their decision was dictated by the nature of studies included for synthesis. For example, Akparibo and others (2017) justified that *'methods of quantitative data synthesis were largely informed by the nature and type of studies identified and included in the review'*. Others, such as Ghose *et al.* (2017) and Hossain *et al.* (2017b), referred to the impossibility of conducting an aggregative synthesis in the absence of suitable studies, opting for a configurative approach instead. Another reason provided for this decision was the heterogeneity of the included studies (Ali *et al.*, 2017; Kumar *et al.*, 2018; Maynard *et al.*, 2017; Williamson *et al.*, 2017). One of the authors made the following explicit acknowledgement to support their methodological choice: *'to differentiate heterogeneity of data between the studies, the textual narration method was adopted, which helped in bringing more clarity to the study contexts'* (Ali *et al.*, 2017). Other review authors opted to justify their synthesis method of choice by referring to the need to appropriately address the review question:

'For synthesising the studies a quantitative approach, in addition to narrative approach, was adopted...In our view this combination will be a better suited approach to address the review question.' (Babu *et al.*, 2017)

The choice of quality assessment tool was either dictated by the heterogeneity of the evidence available to review (Akparibo *et al.*, 2017; Garn *et al.*, 2017), or justified as being the conventional systematic review approach based on the previous professional experience of the reviewers (Yount *et al.*, 2017), or external expert group opinion (Wolf *et al.*, 2018).

Making claims about quality and validity

The three studies which made methodological claims related to quality and validity based their claims on adherence to an established conventional framework (Ali *et al.*, 2017; Babu *et al.*, 2017; Pilkington *et al.*, 2017). Ali (2017) was guided by the Population, Intervention, Control, Outcome (PICO) framework to inform their inclusion criteria and ensure the appropriateness of the selected studies in relation to the research questions posed, which, it was said, ensured the overall quality and validity

of the study in relation to its research aims. Pilkington *et al.* (2017) used evidence summary profiles adapted from the GRADE system in order to develop a presentation format for key evidence which would reflect its consistency and strength.

3.2 Scope of claims

This section explores the scope of the claims in terms of the generalisability or transferability of study findings in the impact evaluations and systematic reviews. We are interested in uncovering whether and how the claims made about their study findings were *generalisable* to any setting or *transferable* to a specific context. The sections are structured as follows: Section 4.2.1 looks at generalisability and transferability in impact evaluations, and Section looks at 4.2.2 generalisability and transferability in systematic reviews.

3.2.1 Generalisability and transferability in impact evaluations

Of 47 impact evaluation studies, 28 discussed generalisability or whether the studies were 'externally valid'. One study appeared to provide specific claims about transferability. Several studies justified their claims about generalisability by describing the **similarity** between the study contexts and new settings (Aker and Ksoll, 2019; Bett *et al.*, 2018; Carew *et al.*, 2020; Gibbs *et al.*, 2020; Li and Liu, 2020; Mvukiyehe and van der Windt, 2020; Pellegrini, 2018; Piper *et al.*, 2018; Berg *et al.*, 2019; Graves *et al.*, 2018). In these studies, context referred to whether the new settings have similar governance and historical backgrounds or a comparable level of infrastructure and human development. Li and Liu's study investigating the impact of the subway investment in Beijing claimed that its results were '*most externally valid in large, dense cities that have sparse subway systems in place and are considering expansion*' (Li and Liu, 2020, p. ii). The authors further stated that the study was useful for developing policy recommendations for other developing countries, such as India, where the environmental context, such as in regard to PM2.5 levels, is similar to China.

Other studies discussed the **representativeness** of the study population, which they claimed made the study findings generalisable to different settings (Asunka *et al.*, 2019; Cocciolo *et al.*, 2020; Edjekumhene *et al.*, 2019; Hirvonen *et al.*, 2017; Parker *et al.*, 2019; Roth *et al.*, 2017; Saboya *et al.*, 2018). Another justification relating to generalisability was that the intervention programmes had been evaluated and

proved to be **transferable** to multiple settings, which made the authors more confident that the programme '*can be successfully implemented in contexts other than Bangladesh and by organisations other than BRAC*' (Bandiera *et al.*, 2017, p. 27).

Similarly, Blattman and others (2018) suggested that the findings of their study might be transferable if similar provisions were implemented in other cities. At the same time, they pointed out that conducting a large-scale RCT in a new setting would be 'a test of external validity of the programme', supporting the notion that a programme that was found to be effective in different settings could enhance generalisability.

In addition, mechanisms underlying the **programme implementation** were also considered when making a claim about generalisability (Berg *et al.*, 2019; Coleman *et al.*, 2019; Pellegrini, 2018). In regard to an RCT examining behavioural responses to information about contaminated drinking water in Ecuador, the authors claimed that the findings could be generalisable to the Ecuadorian population in the Amazonian provinces. The authors acknowledged the role of the mechanism of action when considering the generalisability of the findings to other settings. They explained that changes in the intervention format could affect participant decisions and actions, which might alter the intervention's effectiveness (Berg *et al.*, 2019; Coleman *et al.*, 2019; Pellegrini, 2018). Another study explained that the stakeholder engagement intervention was designed to work directly with communities and used specific tools to ensure the success, and that programmes that '*do not follow these principles may not yield the same positive outcomes...*' (Coleman *et al.*, 2019, pp. 244–289).

Two studies made general observations about the study findings, stating that they might be applicable to similar settings because the intervention was used in a real-life setting, was cost effective and was considered **scalable** (Armand *et al.*, 2019; Rakotonarivo *et al.*, 2017). One study suggested that the intervention might not be applicable to the general rural population as the study found no evidence of its effectiveness (Banerjee, 2018).

3.2.2 Generalisability and transferability in systematic reviews

Among systematic reviews, we identified eight reviews that highlighted the challenges in generalising the review findings to other settings (De Buck *et al.*, 2017; Hossain *et al.*, 2017a; Hossain *et al.*, 2017b; Maynard *et al.*, 2017; Nidhi *et al.*, 2017; Stone *et al.*, 2020; Waddington *et al.*, 2019; Yount *et al.*, 2017). These challenges were mainly related to the **representativeness** of the population and the context of the studies included in the reviews. The heterogeneity and representativeness of the populations

across the included studies were most commonly cited by the review authors as the key aspects when considering generalisability. Maynard and others (2017) identified 16 factors that might help or hinder the implementation of interventions supporting populations' own shelter self-recovery processes following humanitarian crises. The authors claimed that it was not possible to generalise the findings to other settings as the identified factors were context-specific. Another review assessed whether and how the primary studies' authors generalised their findings to the relevant population of the study. However, due to the small sample size of the population in the included studies, the review authors suggested that the review findings were not generalisable to the larger population (Stone *et al.*, 2020). Programme **adaptation** was discussed in one review: Akparibo and others (2017), examining qualitative research to address acute malnutrition in children in humanitarian emergencies, suggested that the findings from the review needed some local adaptation in order to address the context-specific implementation requirements.

While they did not make specific claims about generalisability or transferability, 11 reviews conducted contextual analysis as part of the review approach to explore the transferability of the review findings to specific settings (Ali *et al.*, 2017; Annamalai *et al.*, 2017; Babu *et al.*, 2017; Ghose *et al.*, 2017; Hossain *et al.*, 2017a; Ilavarasan *et al.*, 2017a; Kumar *et al.*, 2018; Nair *et al.*, 2017a; Nair *et al.*, 2017c; Nidhi *et al.*, 2017). Another seven reviews assessed whether the findings from the included studies were transferable or generalisable to different locations, as part of their quality assessment of the included studies (Alampay *et al.*, 2017; Garn *et al.*, 2017; Hossain *et al.*, 2017a; Hossain *et al.*, 2017b; Kumar *et al.*, 2018; Maynard *et al.*, 2017; Stone *et al.*, 2020).

3.3 Communicating evidence claims

All impact evaluations used a variety of tables to communicate the findings of statistical analyses. For seven studies, these tables were included in the appendices, as opposed to being placed alongside the main text (Amirapu *et al.*, 2020; Bandiera *et al.*, 2017; Brudevold-Newman *et al.*, 2017; Källander *et al.*, 2021; McKenzie, 2017; Najy *et al.*, 2018; Roth *et al.*, 2017). One study, Parker *et al.* (2019), acknowledged that 'some readers prefer a formal table of results', and explicitly provided a visual alternative to communicate claims. Statistical specifications were presented in 18 studies. Eleven studies used geographical maps to display randomisation locations, comparing treatment and control groups. Seven studies presented randomised flowcharts to communicate the randomisation process (Aker and Ksoll, 2019; Armand *et al.*, 2019; Banerjee *et al.*, 2018; Barnett *et al.*, 2018; Cocciolo *et al.*, 2020; Dar *et al.*, 2020; Parker *et al.*, 2019). Four studies presented the data collection or project implementation

timelines (Armand *et al.*, 2019; Avdeenko and Frölich, 2019; Dar *et al.*, 2020; Saboya *et al.*, 2018). Graphs and figures were also common throughout the majority of studies, being used to break down important characteristics of the data.

In systematic reviews, nearly all reviews presented the review processes (inclusion and exclusion criteria, data extraction tools), conceptual frameworks and review findings using tables, diagrams or figures. Flowcharts showing the flow of the study selection were presented in 18 reviews. Eight studies communicated the findings using forest plots, four presented Funnel plots, and one used Egger's test. GRADE evidence statement summary tables were presented in two studies (De Buck *et al.*, 2017; Obuku *et al.*, 2017); two reviews used a case study to present the context of the research findings (Ghose *et al.*, 2017; Hossain *et al.*, 2017b). One review developed a visual online evidence map to communicate with stakeholders, to discuss the scoping review findings, informing the next stage of the review (Langer *et al.*, 2018).

3.4 Framing of reviews to make evidence claims

3.4.1 Framing of reviews

Some reviews were framed by the review team alone, without reporting a clear framing process. This was the case for a variety of review methods: meta-analysis (Alampay *et al.*, 2017; Stone *et al.*, 2020; Wolf *et al.*, 2018); meta-analysis plus (Garn *et al.*, 2017; Waddington *et al.*, 2019); qualitative synthesis (Ali *et al.*, 2017); narrative review (Akparibo *et al.*, 2017; Patel *et al.*, 2017); review of reviews (Ilavarasan *et al.*, 2017a; Ilavarasan *et al.*, 2017b; Nidhi *et al.*, 2017; Obuku *et al.*, 2017; Pilkington *et al.*, 2017). Some review reports made clear that other stakeholders were involved in the framing, again for a variety of review methods: meta-analysis plus (De Buck *et al.*, 2017; Langer *et al.*, 2018; Nair *et al.*, 2017c); narrative review (Catalano Richard *et al.*, 2019; Ghose *et al.*, 2017; Maynard *et al.*, 2017); realist review (Zwi *et al.*, 2018); review of reviews (Annamalai *et al.*, 2017; Menon *et al.*, 2018; Nair *et al.*, 2017a; Nair *et al.*, 2017c).

Some review reports provided additional details, explaining how they were developed. Two framework syntheses identified and developed further frameworks from the literature (Hossain *et al.*, 2017a; Williamson *et al.*, 2017). Two meta-analyses-plus reviews provided details of how stakeholders contributed to their framing (De Buck *et al.*, 2017; Langer *et al.*, 2018). Some narrative reviews provided details of how the framing was developed from the literature alone (Blundo-Canto *et al.*, 2018; Hossain *et*

al., 2017b), and others discussed how the framing was developed both from the literature and with stakeholders (Catalano Richard *et al.*, 2019, Ghose *et al.*, 2017, Maynard *et al.*, 2017). The framing of a realist review was developed from the literature (Zwi *et al.*, 2018), and this was the case for four reviews of reviews (Annamalai *et al.*, 2017; Menon *et al.*, 2018; Nair *et al.*, 2017a, Nair *et al.*, 2017c).

In summary, whether the review was framed with stakeholders or not, and the level of detail reported, did not depend on the choice of review methods. Although the two meta-analyses provided no details about their framing (Alampay *et al.*, 2017; Stone *et al.*, 2020), two meta-analyses-plus did, and incorporated social values into a theory of change (De Buck *et al.*, 2017) or into the design features of interventions analysis (Langer *et al.*, 2018). Similarly, three narrative reviews gave no details of framing (Akparibo *et al.*, 2017; Kumar *et al.*, 2018; Patel *et al.*, 2017), while other narrative reviews incorporated social values into their theory of change (Maynard *et al.*, 2017) or logic model (Catalano Richard *et al.*, 2019). Lastly, although five reviews of reviews offered no details about framing (Babu *et al.*, 2017; Ilavarasan *et al.*, 2017a; Nidhi *et al.*, 2017; Obuku *et al.*, 2017; Pilkington *et al.*, 2017), other reviews of reviews were framed based on social values, such as effectiveness and community participation (Nair *et al.*, 2017b) and solidarity (Alampay *et al.*, 2017; Menon *et al.*, 2018), or used an ecological model to frame the review (Yount *et al.*, 2017) or contextualise the synthesis findings (Nair *et al.*, 2017a; Nair *et al.*, 2017b).

3.4.2 Social values shaping reviews

Social values could be recognised in all of the systematic reviews, although they did not necessarily frame the work. Some social values were those highlighted by Gough *et al.* (2014) as relevant to developing clinical guidelines (e.g. utility, effectiveness, efficiency, justice and equity, solidarity or individualism, participation, transparency and accountability, and sustainability). Some were those that Gough *et al.* (2014) considered less relevant to developing clinical guidelines, such as 'education'. Others were those Gough *et al.* (2014) did not consider at all (e.g. cultural heritage, resilience, market and macroeconomic stability, and economic growth).

Social values were clear in the questions addressed by two reviews: a qualitative synthesis addressing the efficiency and accessibility of informal justice systems (Ali *et al.*, 2017) and a narrative review expressing the social value of solidarity by addressing undernutrition using targets set by the World Health Assembly for 2025 (Menon *et al.*, 2018).

Some review reports provided additional details, explaining how they were developed. Two framework syntheses identified and developed further frameworks from the literature, by combining equity, community engagement and resilience (Hossain *et al.*, 2017b) and child rights, ecological systems theory, vulnerability and resilience (Williamson *et al.*, 2017).

Ecological models framed a meta-analysis plus (Nair *et al.*, 2017c), a narrative review (Hossain *et al.*, 2017b) and a review of reviews (Yount *et al.*, 2017). In two reviews of reviews (Nair *et al.*, 2017a, Nair *et al.*, 2017c), an ecological model was employed to contextualise evidence once it had been synthesised. Less formal contextualisation identified social values (gender inequality, community mobilisation) in a narrative review (Ghose *et al.*, 2017), and other social values (efficiency, participation, cultural norms and traditional beliefs, inequity, vulnerability) in a review of reviews (Nidhi *et al.*, 2017).

Theories of change specifying social values framed the following: a framework synthesis with equity, community engagement and resilience (Hossain *et al.*, 2017a); a meta-analysis plus with the social values of solidarity, rights to water and cultural sensitivity (De Buck *et al.*, 2017); and a narrative review with the social values of efficiency and effectiveness (Maynard *et al.*, 2017). One logic model included the social values of solidarity and individualism (Catalano Richard *et al.*, 2019).

In summary, neither the process of framing (from the literature and/or with stakeholders) nor the structure of framing (ecological model, theory of change, logic model) were associated with particular review methods. However, involving stakeholders in framing the review, and developing a theory of change, logic model or ecological model may have helped to identify social values that contributed to framing the reviews.

3.4.3 Reviews with impact

A 'truth table' that cross-tabulates the different elements of our evidence claims framework (Annex G) reveals how the social impact of reviews – in other words, where they appeared and influenced the world beyond research – was associated with combinations of four sets of attributes:

- stakeholder engagement in debating the focus, scope or backdrop of the review (either for individual reviews or as part of earlier consensus development exercises);
- the integration of social values in the review (as components of the intervention or theory of change, or outcomes of interest, or elsewhere in the text); and

- the certainty or reach of evidence claims (in terms of the quality of the underpinning studies and the transferability of the findings).
- the social impact of reviews, where they appeared and influenced the world beyond research.

Instrumental impact: Tabulating the reviews in this way identified eight reviews that each influenced policy or guidance. All of them featured attributes in each of the four sets. All of them drew on stakeholders' debates. Three reviews were motivated by the Sustainable Development Goals (SDGs), which are the result of a global consensus: SDG 6 on ensuring access to water and sanitation for all (De Buck *et al.*, 2017; Wolf *et al.*, 2018) and SDG 16 on promoting peaceful and inclusive societies for sustainable development, providing access to justice for all and building effective, accountable and inclusive institutions at all levels (Waddington *et al.*, 2019). These three reviews and the five others showing instrumental impact also involved stakeholders in debating the review in order to shape its focus or scope, taking into account the backdrop of social values. The review authors either reported their work being guided by an advisory group (De Buck *et al.*, 2017; Langer *et al.* 2018; Nair *et al.* 2017b; Waddington *et al.*, 2019), by a reference group (Zwi *et al.*, 2018), or by an international NGO funder/manager (De Buck *et al.*, 2017; Garn *et al.*, 2017; Maynard *et al.*, 2017), they reported co-authoring the review (Wolf *et al.*, 2018).

All the reviews with instrumental impact also had social values knitted into the intervention or theory of change (De Buck *et al.* 2017; Garn *et al.*, 2017; Langer *et al.*, 2018; Nair *et al.*, 2017b; Wolf *et al.* 2018) and/or social values that were outcomes of interest (Garn *et al.*, 2017; Maynard *et al.* 2017; Nair *et al.*, 2017b; Waddington *et al.*, 2019; Wolf *et al.*, 2018).

Lastly, all eight reviews were methodologically sound, often meta-analyses, and drew conclusions about how the evidence could be generalised either theoretically or statistically.

Conceptual impact: Two reviews that were not recorded in the public domain as informing policy or guidance nevertheless were cited in documents by international NGOs, showing their contribution to sharing understanding of the issues in policy organisations.

Blundo-Canto *et al.* (2018) systematically reviewed the evidence of the livelihoods impacts of Payments for Environmental Services. Subsequently, the Food and Agriculture Organization of the United Nations drew on their findings for two reports about the SDGs. The earlier report (Durango *et al.*, 2019) noted the finding that although schemes offering payments for environmental services were widely

beneficial in economic terms, few reported improvements in terms of social values, such as local livelihoods, either because they had not been taken into account or because payments were found to be detrimental: for instance, reducing equity in decision-making and cultural values. The later report (Kamowitz, 2020) cites Blundo-Canto *et al.* (2018) and other authors when noting the positive but modest impacts of paying to avoid deforesting, rather than paying for more active management and/or efforts to build social and human capital.

The other review showing conceptual impact, but not instrumental impact, addressed positive youth development programmes in LMICs (Catalano Richard *et al.*, 2019). The findings of this review were briefly included in a WHO Knowledge Summary about women's, children's and adolescents' health (WHO, 2020).

These were both narrative reviews. Neither offered judgements about the quality or transferability of the evidence, but both involved stakeholders as co-authors and integrated social values into the framing of the evidence. The stakeholders involved were affiliated with organisations that collate and support the use of evidence: a government agency provided a co-author for the review addressing payments for environmental services (Blundo-Canto *et al.*, 2018), and evidence-informed international NGOs provided co-authors for the review addressing positive youth development.

Knowledge accessibility: Some reviews that neither informed policy guidance nor enhanced policy organisations' understanding nevertheless appeared to interest some potential users. These reviews had all been uploaded to one or more specialist, searchable repositories of research, even though one of the four sets of attributes was either lacking or limited. Kumar *et al.*'s (2018) review of market-led development approaches had no stakeholders informing the work, although it did offer quality-assured evidence addressing economic and social outcomes, and consideration of the relevance to Nepal. Two narrative reviews in this set (Akparibo *et al.*, 2017; Patel *et al.*, 2017) lacked social values framing the questions and analyses, even though stakeholders were involved. Stone *et al.* (2020) involved no stakeholders in their review, and the consensus conference they cited to emphasise the importance of this area of work had taken place 30 years earlier. Williamson *et al.*'s (2017) review of protection for unaccompanied children following humanitarian crises might be considered an outlier in this group. Despite demonstrating attributes in each of the four sets, it does not appear to be cited in a policy document, only uploaded to several specialist databases.

Sixteen reviews showed no apparent impact in the public domain. Two of these were limited in the certainty or reach of their findings (Nair *et al.*, 2017; li *et al.*, 2017).

Others either lacked stakeholder involvement (Alampay *et al.*, 2017; Hossain *et al.*, 2017b; Hossain *et al.* 2018) or, despite involving stakeholders, did not integrate social values into the questions or analyses (Ghose *et al.*, 2018). The other 10 studies in this set were reviews of reviews that similarly lacked stakeholder involvement (Obuku *et al.*, 2017; Yount *et al.*, 2017) and also lacked integration of social values (Ilavarasan *et al.*, 2017; Ilavarasan *et al.*, 2017a; Menon *et al.*, 2018; Srivastava *et al.*, 2017; Babu *et al.*, 2017; Annamalai *et al.*, 2017; Nair *et al.*, 2017a).

Pathways to impact: In summary, for reviews to have an impact on policy decisions, they require attributes belonging to the first three sets listed above:

- a. stakeholder debate for a single study, OR formal consensus, AND
- b. social values integrated into intervention OR theory of change OR outcomes, AND
- c. rigorous, transferable evidence.

None of these attribute sets alone are sufficient, although the direct involvement of stakeholders, or the integration of social values, may be sufficient to attract the attention of specialist knowledge repositories.

Reviews of reviews, which were typically more limited in all three sets of attributes, are not designed to offer direct evidence for policy documents. Neither was their included evidence disseminated more widely through specialist evidence repositories.

4 Other literature relevant to making and justifying evidence claims in research

This chapter summarises the results of a rapid scoping of relevant literature in other disciplines, such as health, economics, management, computer science and psychology, that explore and evaluate evidence claims and how these claims are justified (Part B). The aim is to provide a broad overview of the current debates relating to making evidence claims and their justification. First, we provide a summary of 45 papers that discuss evidence claims and justification (Section 4.1). Second, we summarise the literature that focuses on the general justification of claims or that interrogates multiple elements in the justification of claims (Section 4.2). The last section presents the key focus of 36 papers that focus on particular aspects of justifying evidence claims (Section 4.3).

4.1 Descriptions of relevant literature making and justifying claims

In summarising the literature that is described in this section, we provide a first step in synthesising a diverse literature about justifying evidence claims in research. The tables provide a descriptive snapshot of the nature of this literature, focusing on 45 studies across 10 discipline areas, published between 1986 and 2022, as shown in Table 2. This includes seven papers published between 1986 and 1987, single papers published in 1995 and 2004, 25 papers published between 2010 and 2019, and 11 published since 2020. The research field was assigned based on the publication domain. Two authors (Cartwright and Gough) have written multiple papers and are listed in different research fields. There are multiple papers from two special issue journals, one from *Evaluation and Program Planning*, dated 1987, focusing on justifying conclusions in naturalistic evaluations, and one from *Educational Psychology Review*, dated 2011, focusing on philosophical, theoretical and methodological considerations of prescriptive statements, along with a follow-up paper (Wecker, 2013).

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Research field published within Total papers (N=45)		Overarching (n=9)	Focus on particular aspects of justifying evidence claims (not mutually exclusive) (n=36)			
		Research design (n=15)	Validity and interpretation (n=24)	Relevance and usefulness (n=10)	Communication (n=8)	
Computer and information sciences	2		1: Mingers and Standing (2020)	1: Mingers and Standing (2020)	1: Denning (2011)	1: Denning (2011)
Design	1		1: Vermaas (2016)			
Economics	1			1: Claveau (2011)		
Education	9	2: Gough (2021), Williams (1986)	3: Marley and Levin (2011), Wecker (2013), Yi and Duval-Couetil (2022)	5: Gorard and Tan (2022), Marley and Levin (2011), Snodgrass <i>et al.</i> (2022), Wecker (2013), Yi and Duval-Couetil (2022)	1: Evers and Mason (2011)	2: Nolen and Talbert (2011), Wecker (2013)
Environment	1	1: Delahais and Toulemonde (2017)				
Healthcare and medicine	7		1: Hilton and Jonas (2017)	5: Avis (1995), Bleakley <i>et al.</i> (2020), Lazarus <i>et al.</i> (2015), Shyagali <i>et al.</i> (2022), Yavchitz <i>et al.</i> (2016)	2: Lazarus <i>et al.</i> (2015), Shyagali <i>et al.</i> (2022)	5: Avis (1995), Boutron and Ravaud (2018), Lazarus <i>et al.</i> (2015), Shyagali <i>et al.</i> (2022), Yavchitz <i>et al.</i> (2016)
International development	3		2: Attanasio and Cavatorta (2017), Zalanga (2011)	1: Zalanga (2011)	1: Cartwright <i>et al.</i> (2020)	
Other social science/ social policy	12	5: Cartwright <i>et al.</i> (2010), Davidson (2014), Farley (1987), Gough and	2: Alsharari and Al-Shboul (2019), Moran-Ellis <i>et al.</i> (2016)	5: Alsharari and Al-Shboul (2019), Chrisp <i>et al.</i> (2022),	3: Chrisp <i>et al.</i> (2022), Greene (1987), McClintock (1987)	

Research field published within Total papers (N=45)		Overarching (n=9)	Focus on particular aspects of justifying evidence claims (not mutually exclusive) (n=36)			
		White (2018), Smith, N. (1987)		Greene (1987), Pearsol (1987), Smith, J. (1987)		
Philosophy	5	1: Peters (2020)	2: Cartwright (2021), Kuorikoski and Marchionni (2016)	3: Cartwright (2021), Morton (2011), Staley (2004)	1: Cartwright (2021)	
Psychology	4		3: Antonakis <i>et al.</i> (2010), Haiyan (2011), Kulikowich and Sperling (2011)	3: Antonakis <i>et al.</i> (2010), Kulikowich and Sperling (2011), Sun and Pan (2011)	1: Sun and Pan (2011)	

Table 2 Sampled literature on justifying evidence claims (N=45)

4.2 Summary of overarching literature on justifying claims (n=9)

The following literature, also described in Table 7 in the supplementary file, sets out challenges, perspectives and frameworks relating to justifying the evidence claims of evaluation research.

Cartwright *et al.* (2010) argues for a theory of evidence to apply in practice that is philosophically grounded, particularly in regard to considering what is relevant evidence and how it can be used to evaluate a hypothesis. They highlight the need for credible auxiliary assumptions underpinning evidence and problems in using evidence-ranking schemes. Davidson (2014) presents a toolkit-style paper which can guide evaluators and commissioners in social policy when planning to present claims from evaluations. This sets out the process of 'evaluative reasoning', which is used to draw conclusions from interpreting evidence and for 'synthesising answers to lower- and mid-level questions into defensible judgements that directly answer high-level questions' (p. 3). The process also involves combining judgements with defined concepts of 'quality' and 'value'. Delahais and Toulemonde (2017) reflect on how causal contribution claims were made based on an evaluation related to sustainable forest management. They generalise this by considering the credibility of causal claims where a theory of change is complex. They use a contribution analysis approach and process tracing framework and conclude that contribution analysis

alone does not address the justification of causal claims, though it can be combined with other analytical approaches to improve the rigour of claims.

Farley (1987) critiques the criteria for justifying conclusions in naturalistic evaluation assumptions and proposes a framework involving three elements: empirical accuracy, practical considerations and evaluative considerations. The paper also considers how different views affect the criteria used to justify evidence claims. Smith, N. (1987) considers the kinds of questions associated with four types of claims (research, policy, evaluation, management) in evaluation research and discusses issues relating to justifying evaluation conclusions.

Gough (2021) sets out a framework for considering the 'fitness for purpose of an evidence claim', based on the trustworthiness and relevance of evidence claims. This paper also argues for a greater emphasis on perspectives in using and producing research in order to improve clarity on the usefulness of evidence claims which compete. The debate references Gough and White (2018), which sets out recommendations for evidence standards used in evidence portals based on a survey of 14 evidence portals, with the purpose of helping research users understand what makes research claims reliable. Gough's framework is adapted within the rapid evidence review by Chrisp *et al.* (2022). Several recommendations acknowledge the impact values have on guidance and evidence claims.

Peters (2020) considers that 'mixed claims in science' refers to generalisations that are informed partly by value judgements. This paper considers how values relate to objectivity.

Williams (1986) explores the practical implications of using naturalistic evaluations and some of the criteria based on which evidence claims can be made. The discussion highlights areas that are helpful and conflicting, and suggests that compromise is needed, involving 'cautious flexibility'. Value interpretations are mentioned throughout and 'Valuational interpretation' is one of the evidence standards described.

4.3 Summary of literature focusing on particular aspects of justifying evidence claims (n=36)

The papers detailed in Table 8 in the supplementary file are described below under the following categories: research design; triangulation; validity and interpretation of research findings; relevance and usefulness; and communication. Inevitably, there is

overlap across these categories, and the headings serve as a broad guide to describe the variety of literature in this area.

4.3.1 Research design and making a claim

Antonakis *et al.* (2010) consider when claims can be made about causal inferences in non-experimental settings within the fields of management and applied psychology. They review methods for testing claims and analyse 110 articles in terms of methodological rigour, then set out 10 best practices for causal analysis and summarise 14 threats to validity. They focus on research design and analysis methods for making valid causal inferences.

Kulikowich and Sperling (2011) introduce a special journal issue on making prescriptive statements. They discuss three difficulties in making justified causal/prescriptive evidence claims within educational psychology research: 1) the measurement of unobserved constructs; 2) the complexity of theoretical frameworks; and 3) statistical modelling procedures. Wecker (2013) extends this work by discussing normativity and generality. Marley and Levin (2011) discuss how educational researchers can produce evidence that is credibly linked to a causal factor, and which conditions justify prescriptive statements. They propose a stage model of educational research: 'CAREfully crafted intervention research', a model that requires clarity of hypotheses, methodological checks, testing competing claims, and the role of replication (Kulikowich and Sperling, 2011). Vermaas (2016) discusses the justification of knowledge claims within the research field of design. This paper focuses on the 'precedence of a design method and the justification of the method'.

Mingers and Standing (2020) is a discussion paper that sets out a framework for research validity, which is tested on a range of research forms, including positivist, interpretive, design science, critical, and action-oriented research. The authors argue that there are fundamental validation criteria that can be applied to all research approaches, despite apparent diversities or conflict, and that this can improve research, provide some commonalities and encourage further development in this area. Yi and Duval-Couetil (2022) propose guidelines for enhancing methodological and reporting practices within the field of entrepreneurship education. They develop a 'descriptive validity framework' for evaluating impact research in entrepreneurship education, drawing on previous research syntheses and research guidelines.

4.3.2 Triangulation

Triangulation can be part of the research design and can also support the validity and interpretation of claims. Cartwright (2021), Kuorikoski and Marchionni (2016) and Claveau (2011) consider the justification of causal inferences and the need for evidential diversity from a mix of methods. Cartwright (2021) focuses on the evidence to support multiple subsidiary claims, and characterise perspectives of rigour in context of claims from RCTs, and their challenges. Kuorikoski and Marchionni (2016) use an example of neuroeconomics evidence to triangulate social preferences and then consider its value in improving rigour and controlling for some errors and bias. Claveau (2011) draws on a case study of institutional determinants of the aggregate unemployment rate to use a variety of evidence to justify causal inferences. Attanasio and Cavatorta (2017) discuss the main barriers to making causal inferences within the context of policy evaluation in international development. They propose alternative and complementary approaches to RCTs and provide examples from sectors where standard RCTs are less likely to be applicable: for example, the effect of economic growth on the onset of civil war, and transport infrastructure and its effect on urbanisation.

Triangulation is also considered for claims other than causal claims. Moran-Ellis *et al.* (2016) discuss the epistemology of triangulation and the underpinning belief that more can be learned about a phenomenon through multiple methods. They debate existing definitions of triangulation, integration etc. to provide greater clarity on what constitutes triangulation and when it is valid, including at different stages of research (data collection, analysis, theorising findings). Alsharari and Al-Shboul (2019) discuss knowledge claims in management accounting research using qualitative methods and set out criteria for claims in terms of authenticity, plausibility and criticality, particularly in regard to findings from interpretive case studies for considering the impact of changes over time. They challenge the notion of 'value-free' research within management accounting. They suggest the need to triangulate the findings from multiple sources of data, supported by checks on validity and consideration of the strengths and weaknesses of individual sources. Zalanga (2011) critically reviews challenges in applying statistical techniques for explanation and reasoning within comparative macro-social science research. The paper discusses the value of narratives in validating knowledge claims within low-income country studies in terms of the methodology for data collection and analysis, and the importance of considering both the local context and broader, diverse socio-cultural and historical contexts. Pearsol (1987) critiques existing criteria, such as trustworthiness, and links interpretive perspectives to naturalistic evaluation and applies interpretive logic to justifying evidence claims. The paper suggests that drawing conclusions depends on the evaluator to persuade and analyse the available perspectives,, evidence and

claims made. Pearson acknowledges the importance and challenges of integrating values into social inquiry and suggests that evaluators must choose among value interpretations.

Hilton and Jonas (2017) provide an overview of a method (Claim Assessment Profile) that can be used to capture evidence in healthcare evaluations, particularly where only anecdotal evidence exists, and to assess if the effectiveness of practices can be evaluated. The Claim Assessment Profile approach aims to: '(1) describe and clarify a claim, (2) gather information about the current practice, and (3) determine the practices capacity to participate in further evaluation'.

4.3.3 Validity and interpretation of research findings for making a claim

Gorard and Tan (2022) discuss three different knowledge claims, the problems involved in making them, and their justification. They consider claims as 'fully descriptive', 'generally descriptive' and 'causal'. They develop and critique a model of a plausible causal claim. For all claims, they point out that data from real-world contexts are always tentative and that making claims requires the application of care and good judgement. Staley (2004) debates the 'evidential value of robustness' and its limitations, with an example from experimental particle physics. Although Staley draws on 'Deborah Mayo's error-statistical theory of evidence', the author aims to generalise to any theory of evidence where 'evidential relations supervene on facts about the reliability of testing or inferential procedures'.

Morton (2011) discusses approaches to reasoning and its conventions. The author explores different theories, critiques the logic, and highlights the subjectivity of truths. Morton acknowledges that normative values underpin all research. Avis (1995) critiques existing validity criteria in healthcare research that are binary and instrumentalist, and considers that a distinction between qualitative and quantitative methods for rigour is misleading. Morton suggests that the strength of empirical evidence should be related to epistemological arguments that underlie research (realist, relativist, empiricist). Sun and Pan (2011) consider rigour in prescriptive statements in quantitative educational psychology research. They discuss philosophical and methodological frameworks and conventions within the context of inductive, abductive and deductive approaches. They make recommendations on undertaking replications and developing research programmes, and on the evidence required to make an appropriate prescriptive statement, and on dissemination in the discussion section of a research article. They state that recommendations can be made for the generalisable population where this is clearly defined.

Greene (1987) establishes criteria for justifying and resolving multiple conclusions within naturalistic evaluations. The author sets out the logic supporting a practical approach, with a focus on credibility and utility, combined with a participatory evaluation process from multiple values. Greene also discusses who the conclusions are useful for, and why. Smith, J. (1987) discusses relativism and justifying conclusions in naturalistic evaluations and critiques four papers within the same journal issue on naturalistic evaluation (Greene, 1987; McClintock, 1987; Pearsol, 1987; Smith, N., 1987). Smith, J. considers relativism to be a value judgement.

Chrisp *et al.* (2022) implement an approach to assess evidence claims as applied to a rapid evidence review of a policy problem (basic income experiments). This is informed by the framework by Gough (Gough, 2021; Gough and White, 2018). They assess evidence claims made and the extent to which they were trustworthy and relevant. This includes consideration of certainty, threats, strengths and balances.

Two papers are particularly context-specific. Bleakley *et al.* (2020) make recommendations for estimating false-positive reporting within high-quality RCTs in the sports physical therapy field. Snodgrass *et al.* (2022) is a literature review that integrates 47 articles containing critiques and recommendations on social validity assessment.

4.3.4 Relevance and usefulness of making claims

Cartwright *et al.* (2020) is a methods working paper that focuses on the reliability and relevance of programme prediction when evaluating programme effectiveness within specific local settings. The authors set out a theory of change named the 'causal-process-tracing theory of change' as an approach to using systematic and informed processes and applies this to three case studies. They classify assumptions for making predictions and evaluations into six categories: overall middle-level theory; mid-level causal principles guiding each step; support factors; derailers; safeguards; and range of application.

Evers and Mason (2011) consider the justification of claims within the context of the culture and methods of reasoning within education research (analytical, enumerative, hypothetico-deductive and abductive reasoning). They argue for applying a comparative perspective in research methodologies.

McClintock (1987) discusses the justification of conclusions in naturalistic evaluations within management, from an administration perspective. The author develops a

conceptual framework describing how administrators function as brokers for various types of information, as a means of developing, maintaining or improving programmes. This is informed by a theoretical framework for formative evaluation previously developed by the author that is based on a combination of methodology, programme theory, and evaluation context, which aims to guide evaluation design and analysis.

Denning (2011) is a practice paper aimed at information technology (IT) professionals and sets out an 'anatomy of a grounded claim' beyond finding data to support conclusions. It states that making and recognising grounding claims can improve the trustworthiness of IT system design. The paper briefly describes the structure of such claims, acceptance criteria for claims, and their support.

In addition, a number of papers already described include aspects of relevance and usefulness in regard to making claims (Cartwright, 2021; Chrisp *et al.*, 2022; Greene, 1987; Sun and Pan, 2011), and there is also overlap with the communication of claims (Lazarus *et al.*, 2015; Shyagali *et al.*, 2022).

4.3.5 Communication of claims

Nolen and Talbert (2011) consider making knowledge claims (prescriptive statements) from qualitative research within educational psychology research. They focus on communication of the claims as 'asserted outcomes', rather than as prescriptive findings, and explain that part of this involves the transparency of data collection, analysis and reflexivity. These asserted outcomes could also potentially guide quantitative research to determine cause and effects.

A number of healthcare papers focus on different types of 'spin', which includes misrepresentation of study findings and which can be either intentional or nonintentional (Yavchitz *et al.*, 2016). The criteria of 'spin' may differ between studies and studies were included within this review where they potentially adapted or developed criteria for 'spin', or where this was unclear. Studies on 'spin' that clearly used an existing checklist were not included in this synthesis. Yavchitz *et al.* (2016) developed a classification of different 39 types of 'spin' in systematic reviews under the three categories of misleading reporting, misleading interpretation, and inappropriate extrapolation, developed from consensus methods. Some types apply to report abstracts and others to full texts. These categories are also adopted by Lazarus *et al.* (2015) in developing a classification of 'spin' in non-randomised studies evaluating an intervention. From a survey of 122 participants, Yavchitz *et al.* (2016) consider the most severe types of 'spin' in abstracts to be recommendations for

clinical practice that are not supported by findings, misleading titles, and selective reporting. Boutron and Ravaud (2018) organise practices of 'spin' under the four headings of misreporting methods, misreporting the results, misinterpretation, and other types of 'spin', informed by a systematic literature search. They set out different types of research 'spin' and consider why researchers 'spin' their reports and how this may affect readers' interpretation, and ways that 'spin' can be reduced. Although these three papers on 'spin' have overlapping authors, they differ in how they describe 'spin', in their approach, and in the context.

A separate study by Shyagali *et al.* (2022) draws on existing literature in order to develop their criteria for 'spin' in abstracts. A further 13 papers were identified (listed in Table 10 in the supplementary file) that also consider criteria for 'spin' or misleading communication within healthcare research. Although they are likely to be similar, there could be some differences in their criteria. Furthermore, Yavchitz *et al.* (2016) suggest that identifying 'spin' is subjective.

4.3.6 Other papers not considered (n=34)

It is useful to briefly compare the 45 sampled papers with the 34 papers identified that are not described above. Table 9 in the supplementary file compares these in terms of research field and whether they focused on claims relating to a particular research design or type of claim. Table 10 in the supplementary file lists 34 papers that were not included in the final sample in terms of the research field and focus of the evidence claims. Only one research field (philology) was not represented in the sample, though only one paper in this field was identified (Wellmar, 1993). While 17 out of 24 available papers in healthcare and medicine were not sampled, 13 these focused on 'spin' in reporting or misleading communication, and are likely to draw out similar ideas. The remaining 10 papers covered six discipline areas and focused on a variety of aspects on evidence claims. Sixteen of the 34 papers had a clear focus on research designs, though all of these were represented in the sample. Ten papers had a clear focus on causal inferences, as compared with 12 papers in the sample.

5 Discussion

5.1 Summary of findings

This review aims to interrogate impact evaluations and systematic reviews in international development to understand the nature of evidence claims and how they are justified, and to identify evidence standards and tools for producing, supporting and assessing these claims. We identified a total of 78 empirical studies, 47 of which are impact evaluations and 31 of which are systematic reviews of effectiveness studies. We also include 45 papers that evaluate or discuss challenges in, and approaches for, making and justifying evidence claims.

5.1.1 Focus and scope of the claims in impact evaluations

Our findings here relate to the impact evaluations included in our review. The majority of these studies were conducted in Sub-Saharan Africa. Nearly half of the 47 impact evaluations are relevant to more than one policy sector. The most common research theme is human development and gender. The studies evaluate a wide range of interventions, from cash transfers to education programmes, governance, social safety nets, or tax policies. Nearly three-quarters discuss the theory of change to guide their research.

Of 47 impact evaluations, 31 studies employ experimental or quasi-experimental research methods and 16 use mixed-methods research. For more than three-quarters of the studies, the first author is from an institution based in a high-income country (78%) and only 4% are based at institutions in low-income countries.

Our findings suggest the following:

- Intervention impacts are often communicated alongside the magnitude of the impact, such as effect sizes. The justifications for these claims are based on the technical quality, which involves aligning with academic conventions to prove the certainty, strengths and consistency of the claims. The technical quality justification is occasionally complemented by justification based on the appropriateness of the study design and research methods to address the research questions.
- The focus of the methodological claims clusters around the choice of study design or research methods, and the validity and quality of the research approach. These claims are justified primarily by considering four aspects of how impact evaluations may address research questions on causality: quantity, quality, consistency, and feasibility and appropriateness of the study design.

- When RCTs are not feasible or appropriate due to ethical concerns, time or cost, quasi-experimental research designs, such as difference-in-difference, are preferred as an alternative research design. When employing mixed-methods study designs, the benefits of obtaining rich data that provide insights and complement findings from quantitative research are often used as a justification.
- Specific alternative research methods to minimise biases at the sampling, data collection and analysis stages are justified when facing challenges during programme implementation or in constructing a counterfactual.
- Claims about the quality and validity of the research are largely and implicitly justified by employing well-established study designs (e.g. RCTs) to address bias, or by triangulating data from different sources. When the studies are conducted in more challenging contexts, such as in humanitarian settings, the study authors make reference to the feasibility and appropriateness to justify their choice of research methods.
- Approximately half of the impact evaluations make claims about the generalisability or external validity of the study findings. The study authors justify these claims by considering the population and context of the study and new settings in terms of representativeness, transferability of the findings to multiple settings, and contextual similarity. Mechanisms underlying programme implementation and scalability are also considered when making claims about generalisability.
- All studies use tables, diagrams or figures to communicate empirical knowledge claims. Randomisation maps and flowcharts are less frequently used to communicate how participants are assigned to treatment or control groups. The majority of the studies acknowledge the strengths and limitations that may undermine or strengthen evidence claims. The strengths reported in the studies include employing robust research designs, addressing bias, or having a large sample size. The commonly cited limitations of impact evaluations are baseline imbalances between intervention and control groups, potential confounders, and small sample sizes.

Table 3: Impact evaluations: summary of findings

Nature of the claims	Technical quality: certainty and strengths	Consistency	Appropriateness
Intervention impacts	<ul style="list-style-type: none"> - Effect size, some with explicit guidance on the interpretation of effect sizes - Mechanisms to assess and address potential bias - Robustness check 	<ul style="list-style-type: none"> - Degree to which (in)consistent with the wider literature and theories 	<ul style="list-style-type: none"> - Ethical research standards

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

	<ul style="list-style-type: none"> - Effective randomisation - Internal validity - Sample sizes - Triangulation - Strengths and limitations 	<ul style="list-style-type: none"> - Comparing the magnitude of findings with other studies - Authors' predictions - Alternative explanations - Triangulation 			
Focus of the claims: methodological claims					
Nature of the claims	Quantity	Quality	Consistency	Feasibility and appropriateness	
Choice of study design (RCTs or other experimental, non-experimental, mixed-methods study design)	<ul style="list-style-type: none"> - Data availability - Sample size 	<ul style="list-style-type: none"> - Expected quality of the study design chosen to address the research questions on causality 		<ul style="list-style-type: none"> - Ethical concerns - Cost - Aim of the evaluation - Duration of the study 	
Choice of study methods and technicality (sampling strategies, data collection and analysis methods)	<ul style="list-style-type: none"> - Sampling methods that encourage participation in research 	<ul style="list-style-type: none"> - Addressing biases and spillover effects - Following standardised methods 	<ul style="list-style-type: none"> - Commonly used in the field - Based on previous work 	<ul style="list-style-type: none"> - Challenges faced during implementation - Adaptable 	
Quality and validity	<ul style="list-style-type: none"> - Best available evidence when conducting in conflict-affected settings - Data availability 	<ul style="list-style-type: none"> - RCTs and counterfactual - Addressing biases - Triangulation - Validity of instruments 		<ul style="list-style-type: none"> - Addressing ethical issues - Alternative approaches - Appropriateness and sensitivity to local contexts 	
Scope of the claims: generalisability and transferability claims					
Nature of the claims	Representativeness	Transferability	Contextual similarity	Programme theory	Scalability
External validity and generalisability	<ul style="list-style-type: none"> - Representativeness of the study population 	<ul style="list-style-type: none"> - Proved to be effective and implemented in multiple settings 	<ul style="list-style-type: none"> - Comparable to new settings in terms of governance, historical background, infrastructure 	<ul style="list-style-type: none"> - Mechanism underlying the programme implementation - Mechanism of action 	<ul style="list-style-type: none"> - Cost effective - Real-life setting

			and human development		
--	--	--	-----------------------	--	--

5.1.2 Focus and scope of the claims in systematic reviews

We included 31 systematic reviews to explore how evidence claims are made and justified. The majority of the systematic reviews include research from LMICs. In eight reviews, the specific geographical scope is South Asia, East Asia, and Latin America and the Caribbean. For nearly half of the studies, their first author is from an institution in a high-income country; less than 5% of first authors are based at institutions in low-income countries. Similar to impact evaluations, nearly half of the 31 systematic reviews are relevant to more than one policy sector. The most common policy sector is health. Half of the systematic reviews develop a programme theory to guide the review approach to developing eligibility criteria, selecting appropriate outcomes and synthesis methods. Of the 31 systematic reviews, 14 combine both qualitative and quantitative research findings, 10 synthesise quantitative data, and seven conduct an overview of systematic reviews. A narrative synthesis approach is employed in 17 reviews, and 12 perform a meta-analysis. Other synthesis methods are framework synthesis, thematic synthesis and realist synthesis. Our findings in regard to the nature of evidence claims and justifications for a body of evidence suggest the following:

- Evidence claims about intervention impacts are justified based on four aspects: quantity, quality, consistency and the coherent focus of the evidence included in the reviews.
- When reporting inconclusive findings, limited evidence, inconsistency and a broad focus of the evidence base are often used to justify the claims.
- When no evidence of impact or statistically insignificant findings, or implementation factors, are reported, a main justification is an insufficient evidence base, or limited availability of the evidence base to confirm the findings.
- Research gaps are identified when evidence is insufficient, judged to be of poor quality, or inconsistent.
- The focus of methodological claims clusters around the choice of the review design and review methods, such as the selection of quality assessment tools and synthesis approaches. The claims are justified by the availability and appropriateness of the included studies in regard to addressing the review questions, consistency, and reviewers' and experts' inputs and values. Claims about the validity and quality of the reviews are primarily justified by compliance with the standardised procedure for addressing the review questions.

- The representativeness of the study population is commonly cited as a challenge in regard to generalising the review findings to other contexts. Another consideration when considering whether the findings are generalisable is the applicability of the interventions to a new context, and the implementation challenges in that new context.

We also consider how systematic reviews bring research findings together and how evidence claims are framed in terms of the technical quality of the execution of review methods, the values of reviewers, and the strengths and limitations of the reviews. Our review suggests the following:

- Most reviews develop the protocol in advance to guide the review process. Reviewers' perspectives inform decisions on the choice of review approach, such as whether to include qualitative data or to bring together 'best available' evidence in the absence of high-quality and expected study designs to address the review questions. Assessing the quality of the included studies is becoming the norm. Social values and perspectives illuminating the topic are acknowledged for their potential influence in shaping the review process and how evidence claims are made.
- Nearly half of the review teams engage with external experts. Their expertise informs the review process and decisions by providing guidance and feedback in relation to defining the review scope, identifying relevant literature, developing a search strategy, drawing conclusions and implications, and disseminating review findings.
- The comprehensiveness and robustness of the review process are often cited as the strengths of systematic reviews in regard to supporting evidence claims. This includes having quality assessment and quality assurance processes or engaging with stakeholders in the reviews. The reviews are also valued when they bring together high-quality or different types of evidence, or when they fill research gaps.
- Key limitations when making evidence claims are the heterogeneity of the evidence base, poor methodologies and the poor reporting quality of included studies, publication bias, and a lack of an appropriate study design to answer the review questions.
- The review process is normally communicated using a variety of tables, diagrams and figures, such as flowcharts showing the study selection process. Two reviews provide the context of the review results using case studies. One uses a visual online interactive map to engage with stakeholders in the review.
- Table 4: Systemtic reviews: summary of findings

Systematic review justifications based on the nature of evidence claims				
Framing of claims:				
Social values: utility, efficiency, equity, solidarity/ individualism, heritage and self-determination, sustainability, transparency and accountability				
Apparent in: logic model, theory of change, choice of interventions and outcomes, recurrent themes, stakeholder engagement, research methods				
Focus of the claims: empirical knowledge claims				
Systematic reviews: Justifications based on the nature of evidence claims for a body of evidence				
Focus of the claims: empirical knowledge claims				
Nature of the claims	Quantity	Quality	Consistency	Focus
Intervention impact: intervention (in)effectiveness	Effect sizes, pooled effect size (large/small), availability (sufficient)	High, moderate quality	(In)consistency	Coherent focus in terms of population, locations, types of interventions
Intervention impact: inconclusive findings	Limited availability		Inconsistency	Broad focus
Intervention impact: no evidence of impact or statistically significant results	Insufficient and limited availability			
Research gaps	Insufficient, limited: research on the context lacking	Adequate quality: more rigorous research needed	Inconsistency: more research needed for further analysis	
Programme implementation	Availability and size of evidence			
Focus of the claims: methodological claims				
Nature of the claims	Quantity	Fit for purpose	Consistency	Internal and external value and influence
Choice of review approach (e.g. mixed-		- Appropriateness to address research questions	- Heterogenous nature of the evidence base	-Policy needs

methods, multiple-stages review approach)				
Choice of review methods (search strategy methods, inclusion criteria, quality assessment tools, synthesis)	- Availability of appropriate studies for synthesis	- Type of study design of the included studies - Appropriateness to address research questions	- Heterogenous nature of the included studies	-Reviewers' knowledge and expertise -External experts' inputs
Quality and validity		- Compliance with standardised procedure, guidelines, or conventional frameworks		
Scope of the claims: generalisability and transferability				
Nature of the claims	Representativeness	Applicability		
Challenges in generalising the review findings	- Heterogeneity of the populations across the included studies -Small sample size	- Context-specific - Implementation challenges		

5.1.3 Framing of systematic reviews and their underpinning social values

Social values are apparent in all the reviews analysed, although they are not necessarily addressed systematically. Neither the process of framing (from the literature and/or with stakeholders), nor the structure of framing (ecological model, theory of change, logic model) are associated with particular review methods. However, involving stakeholders in framing the review, and developing a theory of change, logic model or ecological model, may have helped to identify social values that contributed to framing the reviews.

Involving stakeholders in framing reviews and integrating social values into questions and analyses appears to be associated with greater use of their evidence by policy organisations.

5.2 Strengths and limitations

- This review includes two sets of literature: a) impact evaluations and systematic reviews of impact interventions that address global South issues; and b) academic literature from different disciplines discussing approaches to making and

justifying evidence claims. This interdisciplinary approach provides critical insights into the current debates and research gaps, which can inform future research.

- The review is novel, addressing important impact research questions about evidence claims and how these claims are justified. It builds on the conceptual framework of developing, justifying and communicating claims proposed by the Lessons Learnt paper examining the CEDIL-commissioned projects (Oliver *et al.*, 2023). This current review potentially enhances transparency and enables future research to explore evidence claims and evidence standards for research addressing broader review questions beyond impact.
- We extracted and examined data on the nature of evidence claims and justifications. The data extraction process and synthesis were explorative and subjective in nature. The reviewers in the current review may have different views and perspectives on what constitutes evidence claims and their justifications. The data extracted may be implicitly or explicitly stated by the study authors. However, we established mechanisms to maintain consistency and transparency throughout the review process. We piloted the data extraction tools to ensure all reviewers had a clear understanding of the purposes of the tools, with a guidance note for each section. We set up weekly meetings to reflect on the data extraction and the synthesis approach, discussing any disagreement we may have had.
- We purposively searched and selected impact evaluations and systematic reviews funded by development agencies to include a wide range of policy sectors and types of study design. This was to ensure that our review captured and learned about the nature of evidence claims from a variety of explanatory contexts whilst being manageable within the available time and resources. The decision was informed by the peer review feedback and our evaluation of the pilot searches and coding results, which showed that potentially relevant studies are mostly conducted in health-related sectors.
- The findings from the review describe the nature of evidence claims and the justifications of the included studies. We rely on the data in the published papers reported by the study authors. The evidence claims about empirical findings were extracted primarily from the research papers' abstracts and/or executive summaries. In addition, the review's findings on evidence claims and justifications should be interpreted in light of the fact that the findings and conclusions from the included studies may be shaped and informed by key stakeholders, and contextual and structural factors that are not evident or stated by the study authors. However, the review process was iterative. We read and re-read data

extracted from the included studies. When there was insufficient information or a need to further understand the context of the claims, we referred to the full report and extracted new relevant data from the other sections of the paper. We coded data according to the framework and created new codes as new themes emerged.

- We include literature from other disciplines to track the current debates and evaluation about making and justifying claims. This descriptive, scoping review of other relevant literature in other disciplines uniquely brings together insights from diverse literature from multiple research fields that can be used to inform further work in this area. However, while the diversity is a strength it also introduces potential misrepresentation due to our level of understanding within areas less familiar to the research team. Although one of the co-authors of Chrisp *et al.* (2022) is also a co-author of this report we do not consider that this introduces any conflict.

5.3 Conclusion and recommendations

This review explores how evidence claims are supported by research, looking at impact evaluations and systematic reviews of intervention effectiveness in international development. Whilst all of the evidence is relevant to LMICs, for two-thirds of the studies the first authors are based at an institution in a high-income country; only 4% are based at an institution in a low-income country. The nature of the claims broadly focus on impact, implementation factors, the choice of research designs and methods, the credibility of the research, and knowledge gaps. Justifying these various types of claims involves carefully and logically arguing that evidence supporting these claims is credible (in terms of the quality and availability of evidence), consistent and relevant. At the same time, evidence production may be shaped by internal and external influences. Study authors' perspectives and expertise may inform the focus and the choice of research designs and methods that are appropriate and relevant to local priorities. These influences are particularly important in development contexts, which present several challenges to conducting, accessing, and using relevant generalisable research to inform decision-making. (Oliver *et al.*, 2018).

We discuss the scope of evidence claims regarding the extent to which the research findings are generalisable and transferable to new settings. Justification of the scope of evidence claims in impact evaluations is often supported by considering the representativeness of the study population, the transferability of the findings to multiple settings, contextual similarity, and the programme theory underpinning

research and implementation. Although evidence claims about generalisability are less apparent in systematic reviews, applicability and implementation challenges are often discussed.

This review has identified the following research gaps, implications and recommendations:

- This review is one of the first studies to seek to synthesise research in international development, exploring how evidence claims are made and justified in impact evaluations and systematic reviews of intervention effectiveness. Whilst this advances our understanding of how effectiveness evidence claims are made and justified in the field, research asking questions that go beyond the issue of impact should be further explored. Further research may consider a broader focus of claims about equity, innovation, scale-up and sustainability, and how they are framed and justified. This endeavour can inform the development of appropriate and novel research designs and methods that can produce credible, reliable and relevant evidence that can confidently support these claims.
- Unsurprisingly, most of the research evidence focuses on the technical quality, validity, and availability of evidence to justify the claims. It is less clear on how evidence is justified and supported by considering ethical concerns. The need to balance privacy, protection and transparency should be routinely taken into account when designing evaluations and implementing a programme. This should enhance applicability whilst producing socially responsible research.
- Similarly, evidence claims about sustainability and scale-up are less evident in the research included in this review. Evidence diversity can play an important role in supporting such claims. It brings knowledge generated from various types of research and multiple data sources, strengthening collaboration and pertinence to evidence-informed decisions in international development. Guidance and checklists relating to assessing evidence claims on impact, methodological choice, scope, scale-up and sustainability should be developed and routinely used when conducting research.
- The strength of evidence is widely considered to relate to the rigour of studies and how they have been collated. The analysis of the reviews in this study supports the proposition that the strength of evidence in those terms alone (e.g. rigour) is insufficient to influence decisions. An additional essential element is the importance of the issues addressed in that evidence, as expressed by stakeholders involved in shaping reviews or developing prior consensus about the issues. In summary, pathways to evidence use combine the experience of stakeholders with rigorous evidence that addresses important social issues.

- This analysis of methodological rigour, social values and impact helps to distinguish the strength of evidence to withstand scrutiny from the strength of evidence to influence decisions.

ANNEX A: Fitness for purpose of an evidence claim framework (Gough, 2021)

Rigour, explicitness and accountability, coherence, consistency, appropriate and relevance of methods used to make and interpret an evidence claim

- I. The evidence claim
 1. Perspective underlying the claim (Values, priorities, and theoretical constructs), how these are determined, and their fit with researcher users' needs.
 2. Nature, scope, focus, generalisability, certainty, and alternative explanation for the claim.
- II. Basis for the evidence claim
 1. What were the review methods (for bringing together the relevant evidence base)?
 - i. Technical quality for the execution of the review method
 - ii. Appropriateness of the review method
 - iii. Relevance of focus of how the review method was applied
 2. How were included studies appraised?
 - i. Technical quality of the execution of the methods of the included studies
 - ii. Appropriateness of these research methods
 - iii. Relevance of focus of how these methods were applied (including ethics of the process by which the research was undertaken)
 3. What was the resulting totality of evidence used to make the evidence claim?
 - i. Nature of evidence
 - ii. Extent of evidence
 - iii. Does this evidence justify (warrant) the evidence claim?
- III. Appraisal of evidence claims: evidence standards, tools and guides
 1. What are the evidence standards for appraising the warrant?
 2. What methods, guides and tools are used to apply these standards?
 - i. Do these consider all of the relevant technical issues for the evidence claim being made?
 - ii. Are they themselves technically adequate and appropriate?
 - iii. Are there still dangers of being misled by the narrowness of any appraisal?
- IV. Engaging with evidence

1. Communication: what is the nature of the claims and their warrants?
2. Fitness of purpose: of the evidence claims with the users' needs
3. Recommendations and guidance from evidence
 - i. Is the guidance relevant to the decision?
 - ii. What other information and perspectives were used to develop the guidance?
 - iii. Was the process for making these recommendations rigorous, explicit, accountable and appropriate?

ANNEX B: Data extraction tools

Data extraction tool for impact evaluations

- Section A: Bibliographic details
 - A.1) Publication details
(Please include other publications for communication, e.g. blog, policy brief, website)
 - Peer-reviewed journal article
 - Research report
 - Programme document (e.g. monitoring and evaluation reports)
 - Conferences
 - Website
 - Blog
 - Policy brief
 - Discussion paper
 - Working paper
 - Others, please specify
 - A.2) Country of first author, by considering the institution at which they are based
 - Not stated
 - Please specify
- Section B: Population characteristics
 - B.1) Age group (as reported in the study as part of the inclusion criteria for sampling)
 - No specific age group focus (all or one or more age groups)
 - Children and young people only (0–25 years old)
 - Adults only (as specified in the study)
 - Older people only (as specified in the study)
 - Data from other sources, such as documents and administrative data (please specify)
 - B.2) Any marginalised groups (as reported in the study as part of the inclusion criteria for sampling)
 - Not stated
 - Please specify
 - B.3) Gender focus (as report in the study or as the inclusion criteria for sampling)
 - No specific focus
 - Please specify
 - B.4) Any other socio-demographic (as reported in the study as part of the inclusion criteria for sampling)
 - Not stated

- Please specify
- B.5 Equity dimension
 - Socio-economic status (SES)
 - Sex
 - Age
 - Education
 - Place of residence
 - Disability
 - Not stated
 - Head of household
 - Conflict-affected
 - HIV/AIDS
- Section C: Study characteristics
 - C.1) What are the aims of the study? (as reported in the study)
 - Not stated
 - Please specify
 - C.2) What is the objective of the study?
 - To evaluate the effectiveness of an intervention
 - To evaluate the process of delivery or receipt of participation in an intervention
 - To evaluate/synthesise evidence on mechanisms/contextual factors
 - C.3) In which country/countries was the study carried out? (please specify)
 - Not stated
 - Please specify
 - C.4) Development policy sectors
 - Agriculture, fishing and forestry
 - Education
 - Energy and extractives
 - Banking and financial services
 - Health
 - Business, industry, trade and services
 - Information and communication
 - Public administration
 - Social protection
 - Transportation
 - Water, sanitation, waste
 - Humanitarian and emergencies
 - Environment
 - Governance, civil society, and democracy
 - Gender and social exclusion
 - Multisector
 - C.5) Themes
 - Human development and gender
 - Public sector management

- Social development and protection
- Urban and rural development
- Finance
- Private sector development
- Environment and natural resource management
- Economic policy
- C.6) Was any theory discussed in the study?
 - No information
 - Theory positioning
(Introducing theories either at the beginning or end of a study, but their relevance to the study may be unclear. For example, theory is referred to only in the introduction or in the literature review section and never revisited later in the study (research design or findings or conclusions))
 - Theory application
*(The contribution of the theory to the study is discussed or the theories are used to inform research design and data analysis.
'Theory application refers to when an author identified a particular theory or theories early on in the article and then applied that theory or theories in some capacity' (Kusmasi et al., 2013, p. 179)).*
 - Theory testing/generation
Empirically validating an existing theory or building, revising or expanding a theory
- Section D: Intervention
 - D.1) Was the theory of change or logic model discussed in the study?
 - Not stated
 - Yes, please specify
 - D.2) Was this intervention adapted from a previously evaluated intervention?
 - No
 - Not stated
 - Yes, please specify
 - D.3) Intervention name and description
 - Please provide the name and a short summary of the intervention
- Section E: Research methodology
 - E.1) Study design
 - Quantitative
 - Randomised controlled trials (RCTs)
 - Quasi RCTs
 - Cluster RCTs (cRCTs)
 - Controlled group
 - Field experiment
 - Cross-sectional
 - Quasi-experimental methods
 - Non-random controlled group

- Longitudinal study
 - Qualitative
 - Mixed methods
- E.2) Methods of data collection (please specify based on description in the paper)
 - Not stated
 - Please specify
- E.3) Methods of data analysis (please specify based on description in the paper)
 - Not stated
 - Please specify
- Section F: Evidence claims on: Research gaps (GR)
 - F.1) Did the authors discuss the research gaps (GR) of the research? If yes, please provide at least ONE and a maximum of THREE evidence claims.
 - *Please identify one main evidence claim on research gaps and the rationale provided to justify the contributions or the importance of conducting the research. The claim could be (but not limited to): What do we (not) know? What is important to know? How the study might contribute to research, policy and practice. – Any discussion on evidence users' needs*
 - No information – move to the next section
 - Yes – GR1: Research gaps/rationale claim ONE
Please provide a quote for the claim, and page numbers
 - Yes – GR2: Research gaps/rationale claim TWO
 - Yes – GR3: Research gaps/rationale claim THREE
 - F.2) Was any evidence standard applied to make the evidence claim? Your answer must correspond/link to the claim statement in question F.1
 - GR1: Evidence standards
 - No information
 - Yes, please specify
 - GR2: Evidence standards (if applicable)
 - No information
 - Yes, please specify
 - GR3: Evidence standards (if applicable)
 - No information
 - Yes, please specify
 - F.3) Was any guideline/tool/framework used to develop or justify research gaps/rationale? Your answer must correspond/link to the claim statement in question F.1
 - GR1: Evidence tools/guideline/framework
 - No information
 - Yes, please specify
 - GR2: Evidence tools/guideline/framework, if applicable

- No information
 - Yes, please specify
 - GR3: Evidence tools/guideline/framework, if applicable
 - No information
 - Yes, please specify
- F.4) Did the authors use any other outputs, standardised text, template, and/or image or diagram to communicate the claim? Your answer must correspond/link to the evidence claim statement in question F.1
 - GR1: Communicating evidence
 - No information
 - Yes, please specify
 - GR2: Communicating evidence, if applicable
 - No information
 - Yes, please specify
 - GR3: Communicating evidence, if applicable
 - No information
 - Yes, please specify
- Section G: Evidence claims on: research methodology (RM)
 - G.1) Did the authors discuss the appropriateness or technical quality of the chosen research approaches of the impact evaluation? If yes, please provide at least ONE and a maximum of THREE evidence claims.
 - No information – move to the next section
 - Yes- RM1: Research method claim ONE
 - Yes- RM2: Research method claim TWO
 - Yes- RM3: Research method claim THREE
 - G.2) Was any evidence standard applied to make the evidence claim? Your answer must correspond/link to the claim statement in question G.1
 - RM1: Evidence standards
 - No information
 - Yes, please specify
 - RM2: Evidence standards (if applicable)
 - No information
 - Yes, please specify
 - RM3: Evidence standards (if applicable)
 - No information
 - Yes, please specify
 - G.3) Was any guideline/tool/framework used to develop or justify the appropriateness (choice) or technical quality (robustness) of the study? Your answer must correspond/link to the claim statement in question G.1
 - RM1: Evidence tools/guideline/framework
 - No information
 - Yes, please specify
 - RM2: Evidence tools/guideline/framework (if applicable)
 - No information

- Yes, please specify
 - RM3: Evidence tools/guideline/framework (if applicable)
 - No information
 - Yes, please specify
- G.4) Did the authors use any other outputs, standardised text, template, and/or image or diagram to communicate the claim? Your answer must correspond/link to the evidence claim statement in question G.1
 - RM1: Communicating evidence claim
 - No information
 - Yes, please specify
 - RM2: Communicating evidence claim, if applicable
 - No information
 - Yes, please specify
 - RM3: Communicating evidence claim, if applicable
 - No information
 - Yes, please specify
- Section H: Evidence claims on: empirical findings (EF)
 - H.1) What are the main findings from the study?
 - Please specify – providing the statement in the abstract, including quantitative findings (e.g. effect sizes)
 - H.2) Did the authors consider the certainty, impact, strength and consistency of the claim stated in H.1 by considering the appropriateness or technical quality of the research to justify the claim?
 - Unclear
 - Yes, please specify – and provide information on the guideline/tool, if any, that was applied to make/justify the claim on appropriateness or technical quality
 - H.3) Did the authors consider the strengths and limitations of the research that might undermine or strengthen the claim?
 - Unclear
 - Yes, please specify
 - Strengths
 - Limitations
 - H.4) Were any other evidence standards applied to make, justify or communicate claims from empirical findings?
 - No
 - Yes, please specify
- Section I: Evidence claims on: generalisability and transferability (G)
 - I.1) Did the authors discuss the extent to which the results can be implemented in other contexts?
 - No information
 - Yes, please specify
 - I.2) Were any other evidence standards applied to make or justify claims on generalisability and transferability (relating to I.1)?

- No information
 - Yes, please specify
- I.3) Was any guideline or tool applied to justify the claim on generalisability and transferability (relating to I.1)?
 - No information
 - Yes, please specify
- I.4) Did the authors use any other outputs, standardised text, template, and/or image, diagram to communicate the claim (relating to I.1)?
 - No information
 - Yes, please specify
- Section J: Relevance and contribution to potential users
 - No information
 - Please specify

Data extraction tool for systematic reviews

- Section A: Bibliographic details
 - A.1) Publication details
 - Peer-reviewed journal article
 - Research report
 - Conferences
 - Website
 - Blog
 - Policy brief
 - Others, please specify
 - A.2) Country of first author
 - Not stated
 - Please specify
- Section B: Population characteristics
 - B.1) Age group (as reported in the study as part of the inclusion criteria for reviewing)
 - No specific age group focus (all or one or more age groups)
 - Children and young people only (0–25 years old)
 - Adults only (as specified in the study)
 - Older people only (as specified in the study)
 - Data from other sources such as documents, or administrative data – please specify
 - B.2) Any marginalised groups as a focus of the review (as reported in the study as part of the inclusion criteria for reviewing)
 - Not stated
 - Please specify
 - B.3) Gender focus (as report in the study or as part of the inclusion criteria for reviewing)
 - No specific focus
 - Please specify
 - B.4) Any other socio-demographic (as reported in the study as part of the inclusion criteria for reviewing)
 - Not stated
 - Please specify
 - Data from other sources, such as documents or administrative data – please specify
- Section C: Study characteristics
 - C.1) What are the aims of the study or research questions? (As reported in the review)
 - Not stated
 - Please specify
 - C.2) What is the objective of the review?

- To assess the effectiveness of an intervention
- To understand the delivery or receipt of participation in an intervention
- To synthesise evidence on mechanisms/contextual factors
- Effectiveness meta-review
- C.3) What is the geographical scope of the review? (As reported)
 - Global/worldwide
 - LMICs
 - High-income countries
 - Humanitarian emergencies
 - South Asia
 - Southeast Asia
 - Fragile and post-conflict states
 - Latin America and Caribbean
- C.4) Development policy sectors
 - Agriculture, fishing and forestry
 - Education
 - Energy and extractives
 - Banking and financial services
 - Health
 - Business, industry, trade and services
 - Information and communication
 - Public administration
 - Social protection
 - Transportation
 - Water, sanitation, waste
 - Humanitarian emergencies
 - Environment
 - Governance, civil society and democracy
 - Gender and social exclusion
 - Multisector
- C.5) Themes
 - Economic policy
 - Private sector development
 - Finance
 - Human development and gender
 - Social development and protection
 - Urban and rural development
 - Environment and natural resource management
- C.6) Was any theory discussed in the review? (If about the programme theory or logic model of the intervention, please answer in the intervention section)
 - No information
 - Theory positioning

- Theory application
- Theory testing/generation
- C.7) Do not code yet: How was the study framed? What were its underlying values, alternatives, and interpretations?
 - Framing and values of study were apparent, please specify
 - Framing and values of study were unclear
- Section D: Intervention
 - D.1) Was the theory of change or logic model discussed in the review?
 - Not stated
 - Yes, please specify
 - D.2) Intervention name and description
 - Please provide the name and a short summary of the intervention
- Section E: Research methodology
 - E.1) Review approach
 - Please specify
 - E.2) Methods of synthesis
 - Please specify
 - E.3) Did the reviewers assess the quality of the studies included in the review?
 - Not stated
 - Yes, please provide the name of the tool/framework below
 - E.4) Did the reviewers assess the trustworthiness of the body of the evidence? (Totality of the evidence)
 - No
 - Yes, please provide the name of the tool/framework below
- Section F: Evidence claims on: research gaps and rationale (GR)
 - F.1) Did the authors discuss the research gaps and/or rationale (GR) of the research? If yes, please provide at least ONE and a maximum of THREE evidence claims
 - No information
 - Yes – GR1: Research gaps/rationale claim ONE
 - Yes – GR2: Research gaps/rationale claim TWO
 - Yes – GR3: Research gaps/rationale claim THREE
 - F.2) Was any evidence standard applied to make the evidence claim? Your answer must correspond/link to the claim statement in question F.1
 - GR1: Evidence standards
 - No information
 - Yes, please specify
 - GR2: Evidence standards (if applicable)
 - No information
 - Yes, please specify
 - GR3: Evidence standards (if applicable)
 - No information
 - Yes, please specify

- F.3) Was any guideline/tool/framework used to develop or justify research gaps/rationale? Your answer must correspond/link to the claim statement in question F.1
 - GR1: Evidence tools/guideline/framework
 - No information
 - Yes, please specify
 - GR2: Evidence tools/guideline/framework, if applicable
 - No information
 - Yes, please specify
 - GR3: Evidence tools/guideline/framework, if applicable
 - No information
 - Yes, please specify
- F.4) Did the authors use any other outputs, standardised text, template, and/or image or diagram to communicate the claim? Your answer must correspond/link to the evidence claim statement in question F.1
 - GR1: Communicating evidence
 - No information
 - Yes, please specify
 - GR2: Communicating evidence, if applicable
 - No information
 - Yes, please specify
 - GR3: Communicating evidence, if applicable
 - No information
 - Yes, please specify
- Section G: Evidence claims on: research methodology (RM)
 - G.1) Did the authors discuss the appropriateness or technical quality of the chosen review approaches? If yes, please provide at least ONE and a maximum of THREE evidence claims.
 - No information
 - Yes – RM1: Research method claim ONE
 - Yes – RM2: Research method claim TWO
 - Yes – RM3: Research method claim THREE
 - G.2) Was any evidence standard applied to make the evidence claim? Your answer must correspond/link to the claim statement in question G.1
 - RM1: Evidence standards
 - No information
 - Yes, please specify
 - RM2: Evidence standards (if applicable)
 - No information
 - Yes, please specify
 - RM3: Evidence standards (if applicable)
 - No information
 - Yes, please specify

- G.3) Was any guideline/tool/framework used to develop or justify the appropriateness (choice) or technical quality (robustness) of the study? Your answer must correspond/link to the claim statement in question G.1
 - RM1: Evidence tools/guideline/framework
 - No information
 - Yes, please specify
 - RM2: Evidence tools/guideline/framework (if applicable)
 - No information
 - Yes, please specify
 - RM3: Evidence tools/guideline/framework (if applicable)
 - No information
 - Yes, please specify
- G.4) Did the authors use any other outputs, standardised text, template, and/or image or diagram to communicate the claim? Your answer must correspond/link to the evidence claim statement in question G.1
 - RM1: Communicating evidence
 - No information
 - Yes, please specify
 - RM2: Communicating evidence, if applicable
 - No information
 - Yes, please specify
 - RM3: Communicating evidence, if applicable
 - No information
 - Yes, please specify
- Section H: Evidence claims on: empirical findings (EF)
 - H.1) What are the main findings from the review?
 - Please specify – providing a statement, including quantitative findings (e.g. effect sizes)
 - H.2) Did the authors consider the certainty, impact, strength and consistency of the claim stated in H.1 by considering the appropriateness and/or technical quality of the research to justify the claim?
 - Unclear
 - Yes, please specify – and provide information on the guideline/tool, if any, that was applied to make/justify the claim about appropriateness or technical quality
 - H.3) Did the authors consider the strengths and limitations of the research that might undermine or strengthen the claim?
 - Unclear
 - Yes, please specify
 - Strengths
 - Limitations
 - H.4) Were any other evidence standards applied to make, justify or communicate claims from empirical findings?

- No
- Yes, please specify
- Section I: Evidence claims on: generalisability and transferability (G)
 - I.1) Did the authors discuss the extent to which the results can be implemented in other contexts – wider populations (generalisability), or the extent to which the impact of the programme can be achieved in another setting (transferability)?
 - No information
 - Yes, please specify
 - I.2) Were any other evidence standards applied to make or justify claims on generalisability and transferability (relating to I.1)?
 - No information
 - Yes, please specify
 - I.3) Was any guideline or tool applied to justify the claim on generalisability and transferability (relating to I.1)?
 - No information
 - Yes, please specify
 - I.4) Did the authors use any other outputs, standardised text, template, and/or image or diagram to communicate the claim (relating to I.1)?
 - No information
 - Yes, please specify
- Section J: Relevance and contribution to potential users
 - No information
 - Please specify

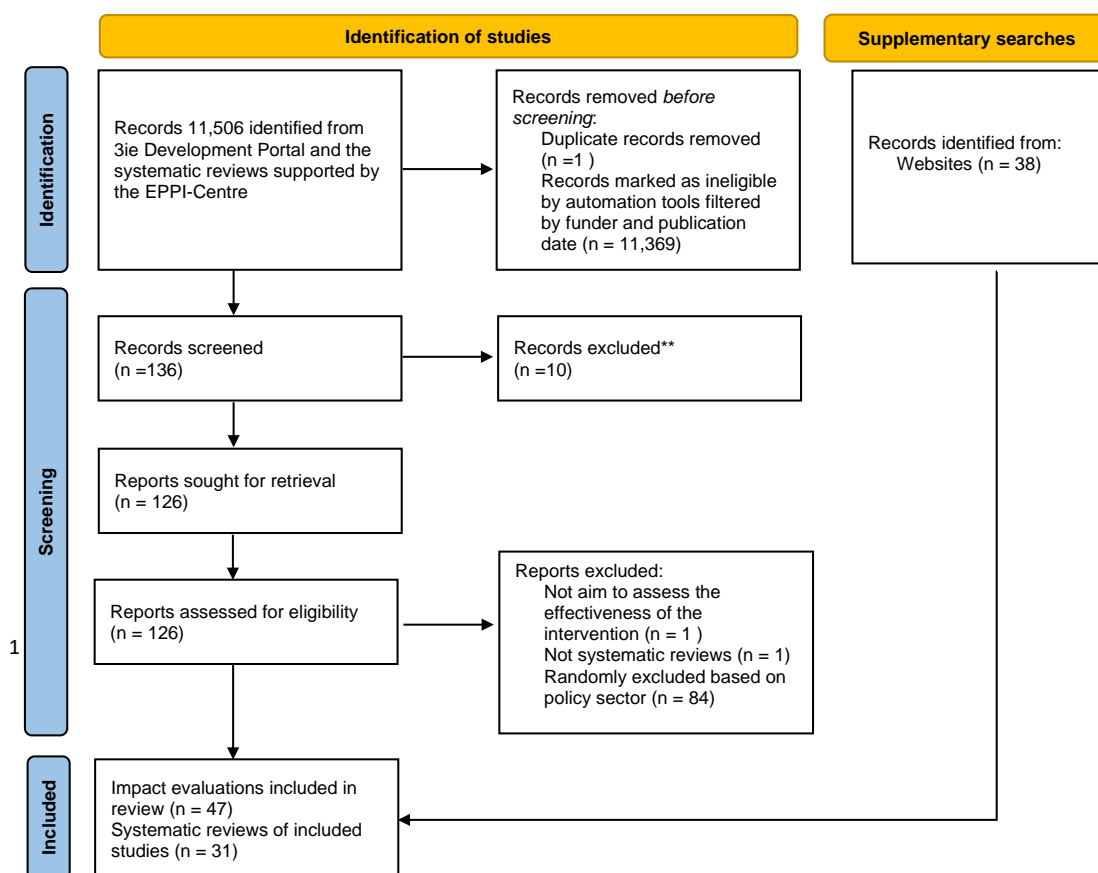
ANNEX C: Findings: a review of impact evaluations and systematic reviews

This section presents the findings of the current review (Part A). We seek to understand the nature of evidence claims in impact evaluations and systematic reviews in international development, and we explore how these claims are developed, justified, and communicated. In this section, we present the search results and the key characteristics of the included studies.

Key characteristics of the impact evaluations and systematic reviews included in the review

In Part A, we identified 11,506 relevant citations from the 3ie Development Evidence Portal and the systematic reviews supported by the EPPI-Centre. After excluding studies using the search function to filter by funder and publication date, 136 studies were included for full-text screening and imported into the systematic review software, EPPI-Reviewer Web (Thomas *et al.*, 2020). After applying full-text screening and an initial coding exercise, a total of 78 studies (47 impact evaluations and 31 systematic reviews) were randomly selected from 126 identified studies for inclusion in this review (see Figure 5). Of the excluded 84 studies, the most common policy sectors covered were health (n=26), social protection (n=15), and agriculture, fishing and forestry (n=12). Other sectors were gender and social exclusion (n=9), environment (n=8), water, sanitation, waste (n=6), education (n=5), humanitarian and emergencies (n=4), information and communication (n=4), transportation (n=4), business, industry and services (n=2), public administration (n=2), energy and extractions (n=1), and governance, civil society and democracy (n=1).

Figure 5: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram



For Part B, a total of 276 papers were identified, of which 14 were duplicates. Of these papers, 78 were identified as relevant after checking the full text, and 45 remained following sampling (see Table 5).

Total records identified, N=276 (including duplicates)	No. of records screened (n=264) (duplicates not counted)	Screened as relevant after checking full-text (n=78)	Sampled includes (n=45)
1) Appendix D of protocol (Bangpan <i>et al.</i> , 2022)	44	11	10

Total records identified, N=276 (including duplicates)	No. of records screened (n=264) (duplicates not counted)	Screened as relevant after checking full-text (n=78)	Sampled includes (n=45)
2) Exploratory searches part 1	75	36	16
3) Exploratory searches part 2, including OpenAlex searches	145, of which: 68 from Open Alex title searches 48 from related papers citation searches 29 from follow-ups after screening papers or <i>ad-hoc</i> processes	31	19

Table 5 Overview of study identification for Part B and their contribution to the final sample of literature

Key characteristics of impact evaluations

The following paragraphs summarise the key characteristics of the 47 impact evaluations.

Population characteristics

The majority of studies did not specify an age group (n=23) or they obtained data from other sources (e.g. documents, administrative data etc). Where age was reported, most of these studies focused on adults (n=15), with far fewer studies focusing on children and young people (n=5). Other population characteristics included studies that focused specifically on women (n=6), poverty (n=10), families with young children (n=2), or other forms of economic and social vulnerabilities (n=7).

Country

Studies were carried out across 27 countries, with the majority conducted in Sub-Saharan Africa (n=16). Other regions included South Asia (n=5), South America (n=3), Southeast Asia (n=2) and East Asia (n=1). The most common countries across all

regions were Bangladesh (n=5) and Uganda (n=5). These were followed by Ethiopia (n=3), Ghana (n=3), Kenya (n=3), Cambodia (n=2), India (n=2), Malawi (n=2), Mozambique (n=2), Niger (n=2) and Pakistan (n=2). The remaining 16 studies took place in Afghanistan, Benin, Chad, China, Colombia, Democratic Republic of the Congo, Ecuador, Lesotho, Madagascar, Nepal, Nigeria, Peru, Philippines, Tanzania, Zambia and Zimbabwe.

When considering the country of the first author's research institution, 79% of the first authors were based at institutions in high-income countries (n=37), 17% at institutions in middle-income countries (n=8) and only 4% at institutions in low-income countries (n=2).

Policy sectors

The included studies fall within 15 development policy sectors, with approximately 43% (n=20) belonging to more than one sector. The most common policy sectors were social protection (n=12), health (n=8), governance, civil society and democracy (n=8) and public administration (n=6). Other sectors were agriculture, fishing and forestry (n=5), gender and social exclusion (n=5), education (n=4), environment (n=4), energy and extractives (n=3), banking and financial services (n=3), information and communication (n=3), business, industry, trade and services (n=2), transportation (n=2), water, sanitation, waste (n=2), and humanitarian and emergencies (n=2). See Figure 6 for types of policy sectors covered, by region.

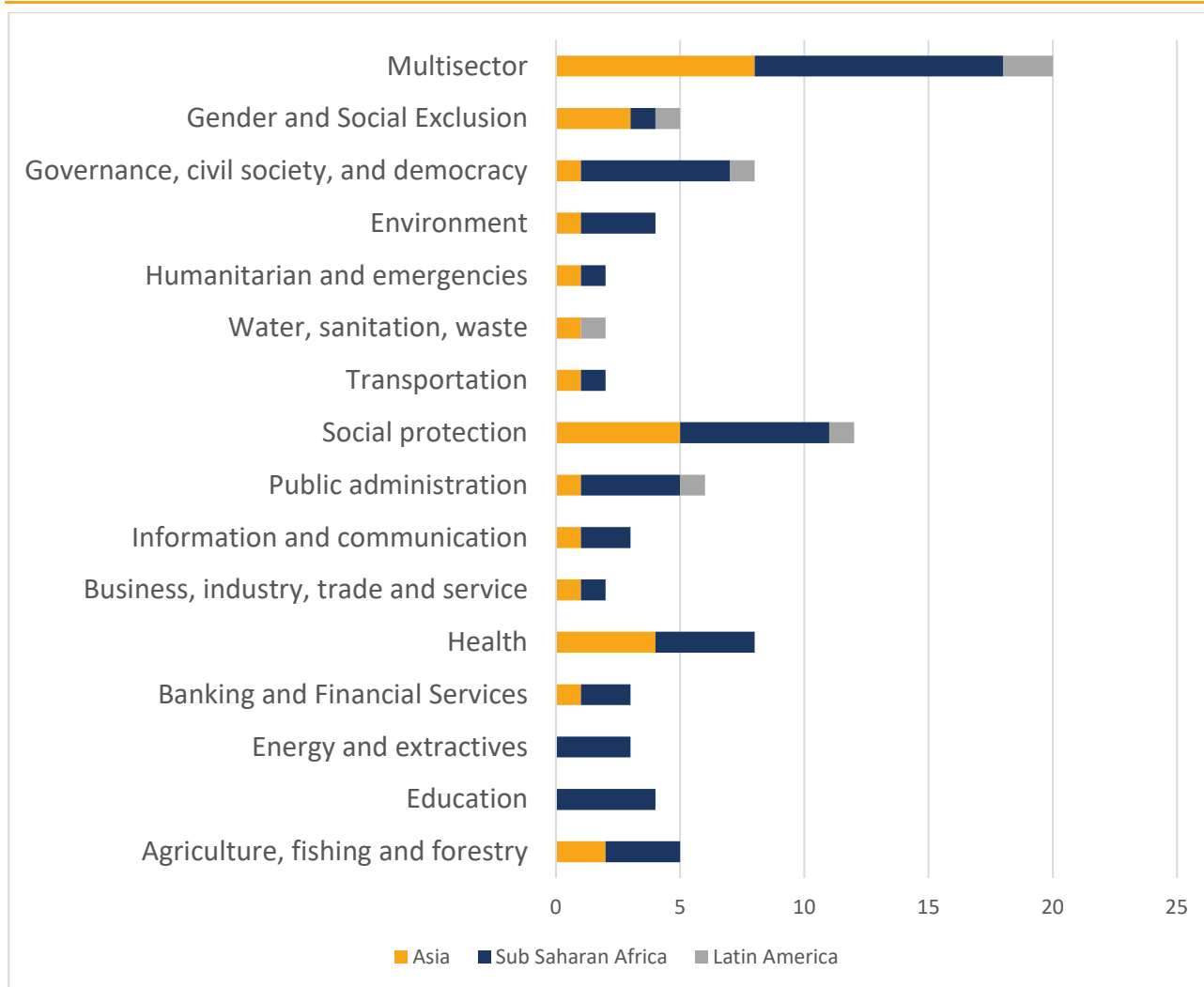


Figure 6: Type of policy sectors covered by impact evaluations, by region (codes not mutually exclusive)

Theoretical framing, programme theory, and intervention

Most of the studies (64%) applied theories or developed theoretical frames to guide, explain and interpret their research findings (n=27). Two studies briefly discussed the nexus of economic theories to support claims on the importance of political accountability for development and the role of entrepreneurship programmes in labour market participation, but there was no assertion of how the theoretical perspectives would be applied to guide the studies (Aker *et al.*, 2017; Brudevold-Newman *et al.*, 2017). One study assessed whether the Discrete Choice Experiment was theoretically valid in regard to predicting the outcomes of compensation schemes for forest conservation (Rakotonarivo *et al.*, 2017). In addition, most of the impact evaluations discussed the theory of change or logic model of the interventions (n=32).

Studies were categorised into eight development themes. The most common themes were human development and gender (n=25), social development and protection (n=15), and urban and rural development (n=11). The remaining themes were public sector management (n=7), environment and natural resource management (n=5), finance (n=5), economic policy (n=3), and private sector development (n=3). Within these themes, there were a wide range of interventions: for example, cash transfer programmes, an adult education programme, reproductive and maternal health interventions, transparency interventions, accountability interventions, governance interventions, social safety nets, rural markets interventions, water pollution interventions, a drought insurance experiment, and tax policy interventions.

For a full description of all studies, see Annex D.

Research methodology

The studies were either purely quantitative by design (n=31) or adopted a mixed-methods approach (n=16). The quantitative designs used included RCTs (n=11), cluster RCTs (n=18), quasi-experimental methods (n=20), field experiments (n=2) and a longitudinal study (n=1) (see Figure 7).

Surveys (n=40) and questionnaires (n=9) were the most common form of data collection, followed by direct measurement by the research team (n=8), the use of secondary or administrative data (n=6), administered assessments (n=1) and observation (n=2). One study also used photographs as part of its survey design. Of the mixed-methods studies, almost all (n=13) used interviews. Other qualitative data collection included focus groups (n=5) and participatory or peer methods (n=1).

All 47 studies contained a quantitative element and conducted statistical analysis. Of the studies that also used qualitative methods, analysis approaches included thematic analysis (n=3), content analysis (n=2) and framework analysis (n=2). Nine of these studies did not define the specific type of qualitative analysis.

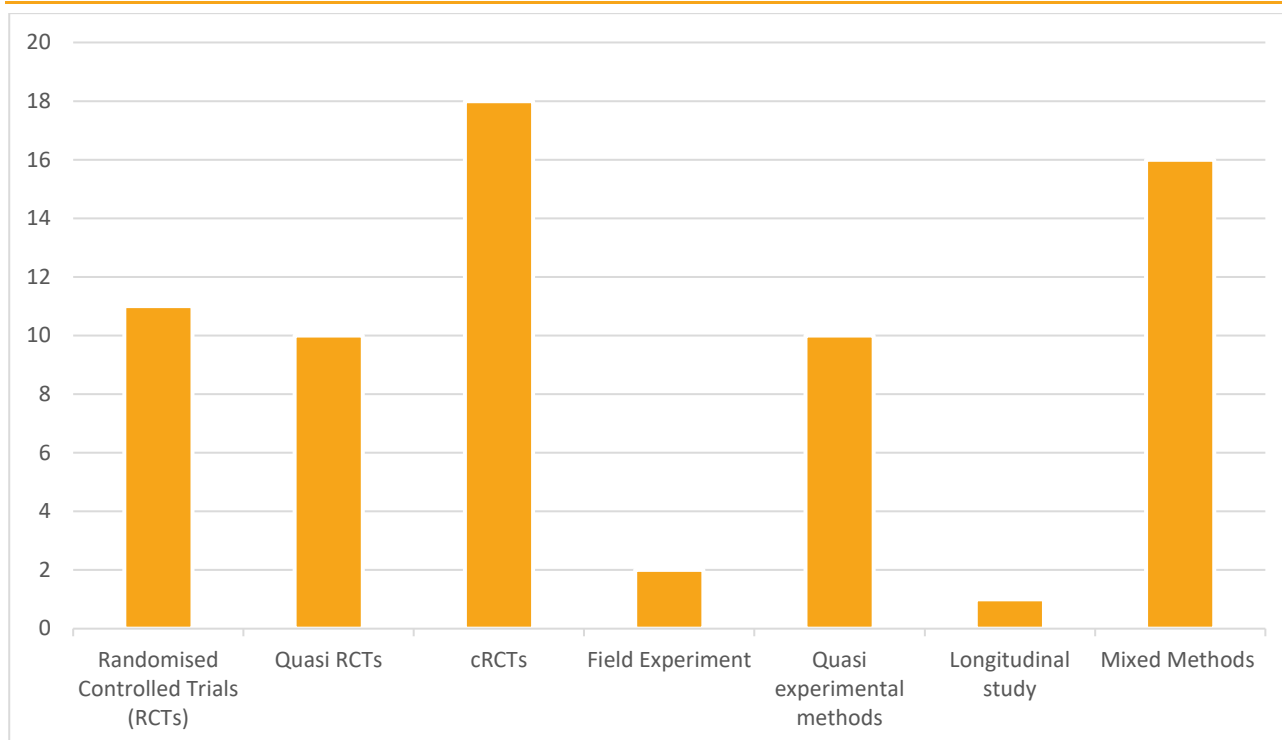


Figure 7: Study design (codes not mutually exclusive)

Key characteristics of systematic reviews

The following paragraphs summarise the key characteristics of the 31 systematic reviews.

Population characteristics

The majority of reviews did not specify an age group (n=24). Of the reviews which reported age, the focus was mostly on children and young people (n=6), with one review specifying adults only. Other population characteristics included reviews focused specifically on women (n=5), urban populations (n=2), rural populations (n=2), migrants (n=1), populations with health conditions (n=1), populations affected by humanitarian crises (n=1), separated children (n=1), those with experience of political instability (n=1), those with vulnerability (n=1) and pregnant women (n=1).

Geographical scope

For most reviews, the geographical scope was defined as LMICs (n=22). For reviews with a specific regional focus, South Asia was the most common (n=6), followed by East Asia (n=1) and Latin America and the Caribbean (n=1). The geographical scope of four reviews was categorised as global/worldwide. Other scopes were defined as countries experiencing humanitarian emergencies (n=2) and fragile and post-conflict states (n=1).

When considering the country of the first author’s research institution, 48% of the first authors were based at an institution in a high-income country (n=15), 48% at an institution in a middle-income country (n=15) and only 4% at an institution in a low-income country (n=1)

Policy sectors

The included reviews fall across 15 development policy sectors, with approximately 42% (n=13) belonging to more than one sector. The most common policy sectors were health (n=7), social protection (n=5), humanitarian emergencies (n=5), business, industry, trade and services (n=4) and water, sanitation and waste (n=4). Other sectors were agriculture, fishing and forestry (n=2), education (n=3), information and communication (n=3), gender and social exclusion (n=2), banking and financial services (n=2), environment (n=2), governance, civil society, and democracy (n=2), energy and extractives (n=1), public administration (n=1) and transportation (n=1). Several reviews had multiple focuses, reviewing global evidence and subsequently conducting a contextual analysis for specific regions or countries. Figure 8 presents the type of policy sectors by income level and geographical focus.

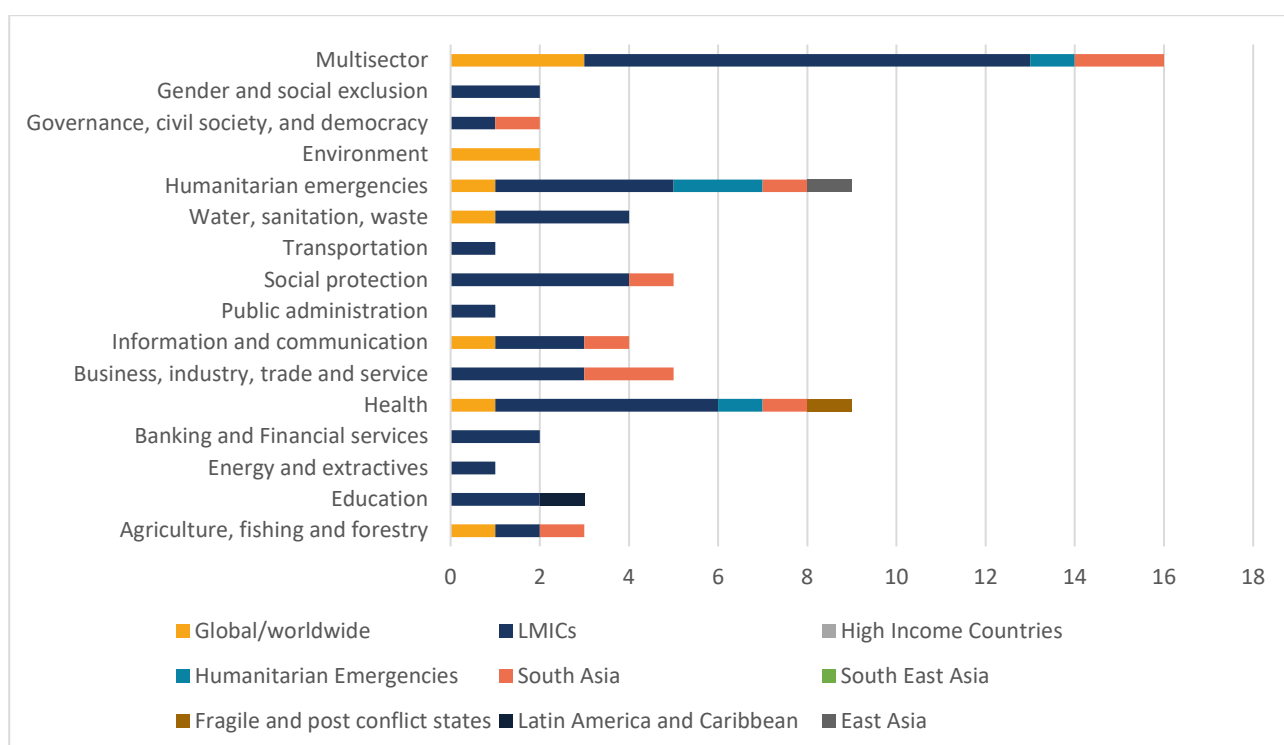


Figure 8: Systematic reviews and type of policy sectors by income level and geographical focus (codes not mutually exclusive)

Theoretical framing, programme theory, and intervention

Over half of the systematic reviews discussed theoretical and conceptual frameworks guiding the review approach for identifying relevant studies or informing decisions on organising and synthesising evidence (n=22). In some cases this included a detailed discussion of relevant theories, such as ecological systems theory (Yount *et al.*, 2017; Williamson *et al.*, 2017; Nair *et al.*, 2017a), theoretical models on growth in developing economies (Babu *et al.*, 2017), and developmental housing theory (Maynard *et al.*, 2017). The majority of the reviews discussed the programme theory used to guide their review approach (n=16), such as developing eligibility criteria, and informing decisions on outcome measures and synthesis. There were a range of interventions (such as mobile financial services, sanitation interventions and citizen engagement interventions), and these focused on seven different development themes. The most common themes were human development and gender (n=12), social development and protection (n=9) and urban and rural development (n=8). The remaining themes were finance (n=5), environment and natural resource management (n=2), private sector development (n=2) and public sector management (n=1).

For a full description of all studies see Annex D.

Research methodology

Just under half of the reviews adopted a mixed-methods approach (n=17), by asking more than one question and/or drawing on multiple types of data to address a single objective. Seven reviews conducted a systematic review of reviews. Six reviews conducted a synthesis of quantitative data, four of which conducted a meta-analysis, and six conducted a numerical narrative synthesis, as meta-analysis was not possible. One review conducting a qualitative evidence synthesis of people's views on receiving or implementing an intervention used thematic methods to analyse the data. Contextualisation of the review findings to explore whether the review findings were applicable or transferable to other settings was conducted in 14 studies. Overall, synthesis approaches included meta-analysis (n=12), narrative synthesis (n=17), framework synthesis (n=4), thematic synthesis (n=2), realist synthesis (n=1), and best-fit framework synthesis (n=1).

All reviews except two assessed the methodological quality of the included studies. The most common approach for doing this was to use appraisal tools specifically tailored to the review (n=10). Other reviews opted to use named or existing tools and frameworks. These included the following: a risk of bias tool (n=6), AMSTAR (n=5),

Critical Appraisal Skills Programme (CASP) (n=5), the Waddington and Hombrados (2012) tool (n=3), the Quality Assessment Tool (n=1), the Mixed Methods Appraisal Tool (MMAT) (n=1), the Newcastle-Ottawa Scale (n=2), the Effective Public Health Practice Project quality assessment tool (n=1), the Liverpool Quality Appraisal Tool (n=1) and the Adapted Checklist for Blueprint Program Evaluation (n=10). Only seven reviews explicitly rated the trustworthiness of the body of evidence. Of these seven reviews, the tools used to do so were Grading of Recommendations Assessment, Development and Evaluation (GRADE) (n=4), Preferred Reporting Items for Systematic Reviews (PRISMA) (n=1) and Campbell Collaboration standards (n=1).

ANNEX D: Key characteristics of studies

Table A: Key characteristics of impact evaluations

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
Ahmed <i>et al.</i> (2019)	<p>Publication details Discussion paper</p> <p>Country of first author Bangladesh</p>	Bangladesh	<p>Development sectors Health</p> <p>Themes Human development and gender</p>	To implement two, linked RCTs in rural Bangladesh, and to explore the mechanisms underlying the impact	<p>Adults and children</p> <p>Poor households</p> <p>Equity: SES, age</p>	Yes	<p>Study design Mixed methods: cRCTs and interviews</p>
Aker <i>et al.</i> (2017)	<p>Publication details Peer-reviewed journal article</p> <p>Country of first author USA</p>	Mozambique	<p>Development sectors Information and communication Public administration Governance, civil society, and democracy</p> <p>Themes</p>	To assess different forms of voter education during an election in Mozambique	Adults only	Yes	<p>Study design Field experiment</p>

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
			Public sector management Social development and protection				
Aker and Ksoll (2019)	Publication details Peer-reviewed journal article Country of first author USA	Niger	Development sectors Education Information and communication Themes Human development and gender	To implement an RCT of an adult education programme in Niger, which included an additional intervention designed to improve teacher accountability and students' learning	Adults only Illiteracy Equity; sex, education	Yes	Study design cRCTs
Amirapu <i>et al.</i> (2020)	Publication details Discussion paper	Bangladesh	Development sectors Social protection Gender and social exclusion	To assess the impact of a child marriage law on social norms and	Adults only Equity; sex	Yes	Study design cRCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Country of first author UK		Themes Human development and gender	marriage behaviour			
Angeles <i>et al.</i> (2019)	Publication details Research report Country of first author USA	Zimbabwe	Development sectors Social protection Themes Human development and gender Social development and protection	To implement and assess the impact of a social cash transfer intervention in Zimbabwe	Poor households Equity; SES	Yes	Study design Mixed methods; quasi RCT
Armand <i>et al.</i> (2019)	Publication details Programme document Country of first author Spain	Mozambique	Development sectors Energy and extractives Governance, civil society, and democracy Themes Public sector management	To implement and assess the impact of three community-level information interventions in Mozambique	Adults only	Yes	Study design cRCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
Asfaw <i>et al.</i> (2017)	<p>Publication details Peer-reviewed journal article</p> <p>Country of first author Italy</p>	Zambia	<p>Development sectors Social protection</p> <p>Themes Social development and protection</p>	To assess social cash transfer programmes against the negative effect of weather risk on rural households' welfare in Zambia	<p>Adults only</p> <p>Women (primary female caregiver)</p> <p>Households with a child aged less than five years old</p> <p>Equity; sex, age, place of residence, disability</p>	Yes	Study design cRCTs
Asingwire (2019)	<p>Publication details Programme document</p> <p>Country of first author Uganda</p>	Uganda	<p>Development sectors Health</p> <p>Themes Human development and gender</p>	To assess the impact of selected components of the Family Planning Programme of the	<p>Children and young people only (0–25 years old)</p> <p>Equity; age</p>	Yes	Study design Mixed methods: quasi RCTs, cRCTs; interviews/focus groups

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
				Government of Uganda			
Asunka <i>et al.</i> (2019)	Publication details Peer-reviewed journal article Country of first author USA	Ghana	Development sectors Public administration Governance, civil society, and democracy Themes Public sector management	To assess the impact of domestic election observers on electoral fraud and violence	Adults only	Yes	Study design cRCTs
Avdeenko and Frölich (2019)	Publication details Programme document Country of first author Germany	Pakistan	Development sectors Social protection Humanitarian and emergencies Themes Human development and Gender Urban and rural development	To assess the impact of natural disaster preparedness interventions in Pakistan	Not specified	Yes	Study design Mixed methods: cRCTs and interviews

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
Bandiera <i>et al.</i> (2017)	Publication details Peer-reviewed journal article Country of first author UK	Bangladesh	Development sectors Agriculture, fishing and forestry Themes Human development and gender	Implementation of a randomised evaluation of a skills intervention for women in Bangladesh	Women Household wealth ranking Equity; SES, sex, head of household	Yes	Study design cRCTs
Banerjee <i>et al.</i> (2018)	Publication details Peer-reviewed journal article Country of first author USA	India	Development sectors Health Themes Human development and gender Nutrition	Reports on the impact of a potential strategy to address iron deficiency anaemia in rural areas: salt fortified with iron and iodine	Not specified	Yes	Study design cRCTs
Belissa <i>et al.</i> (2019)	Publication details Peer-reviewed	Ethiopia	Development sectors Banking and financial services	To report the results of a drought insurance	Adults only	Yes	Study design RCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	journal article Country of first author Netherlands		Themes Human development and gender Finance	experiment in Ethiopia			
Béné <i>et al.</i> (2020)	Publication details Peer-reviewed journal article Country of first author Colombia	Niger	Development sectors Agriculture, fishing and forestry Environment Governance, civil society, and democracy Themes Urban and rural development Finance	To evaluate the effects of a three-year resilience intervention (the SUR1M project in Niger) on the beneficiaries of the project	Not specified	Yes	Study design Quasi-experimental methods
Berg <i>et al.</i> (2019)	Publication details Peer-reviewed journal article Blog	India	Development sectors Health Social protection Themes	To assess the effectiveness of a pay incentive intervention in India	Rural poor Equity; place of residence	Yes	Study design RCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Country of first author UK		Human development and gender Social development and protection				
Berhane <i>et al.</i> (2017)	Publication details Research report Country of first author USA	Ethiopia	Development sectors Social protection Themes Social development and protection Urban and rural development	To assess the impact of a social cash transfer programme	Not specified	Yes	Study design Mixed methods; quasi-experimental methods; interviews, focus groups
Bett <i>et al.</i> (2018)	Publication details Working paper Country of first author Kenya	Malawi	Development sectors Agriculture, fishing and forestry Themes Urban and rural development	To assess the impact of plant clinic activities in Malawi on tomato productivity and farmer knowledge	Not specified	Yes	Study design Quasi-experimental methods
Blattman <i>et al.</i> (2018)	Publication details	Colombia	Development sectors	To assess the impact of a policing	Not specified	Yes	Study design Mixed methods: qualitative -

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Research report Country of first author USA		Public administration Governance, civil society, and democracy Themes Public sector management Social development and protection	intervention on citizen wellbeing in Colombia			interviews; quantitative – RCTs
Brudevold-Newman <i>et al.</i> (2017)	Publication details Discussion paper Country of first author USA	Kenya	Development sectors Education Gender and social exclusion Themes Human development and gender Finance Private sector development	To implement a randomised evaluation of an entrepreneurship intervention in Kenya	Children and young people only (0–25 years old) Women Poor households Equity; SES	Yes	Study design RCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
Carew <i>et al.</i> (2020)	<p>Publication details Peer-reviewed journal article</p> <p>Country of first author UK</p>	Kenya	<p>Development sectors Education Social protection</p> <p>Themes Human development and gender Social development and protection</p>	To assess the impact of a teaching intervention in Kenya	<p>Adults only</p> <p>Equity; disability</p>	Yes	<p>Study design Quasi-experimental methods</p>
Chirwa <i>et al.</i> (2017)	<p>Publication details Peer-reviewed journal article</p> <p>Country of first author USA</p>	Malawi	<p>Development sectors Social protection</p> <p>Themes Human development and gender Social development and protection</p>	To assess the impact of a lean season food transfer on household food security, diet, and nutrition status of young children during the lean season in Malawi and to understand	<p>Food-insecure village</p> <p>Poor households</p> <p>Equity; SES</p>	Yes	<p>Study design Mixed methods: longitudinal study; quasi experimental methods Interviews</p>

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
				processes through which transfers operated			
Chris et al., (2018)	Publication details Research report Country of first author UK	Ghana	Development sectors Agriculture, fishing and forestry Themes Human development and gender Urban and rural development	To assess the impact of the Millennium Villages Project intervention, against the Millennium Development Goals targets, alongside changes in other outcome variables and spillover effects; and to consider its likely sustainability	Poverty status Equity; SES	Yes	Study design Mixed methods: interviews; quasi-experimental methods
Cocciolo <i>et al.</i> (2020)	Publication details	Bangladesh	Development sectors	To assess the impact of a programme to	Data from other sources, such as	Yes	Study design RCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Research report Country of first author Sweden		Water, sanitation, waste Themes Urban and rural development Environment and natural resource management	provide safe sources of drinking water in rural Bangladesh	documents, administrative data Rural poor Equity; SES, place of residence		Quasi-experimental methods
Coleman <i>et al.</i> (2019)	Publication details Peer-reviewed journal article Country of first author USA	Uganda	Development sectors Public administration Environment Themes Environment and natural resource management	To assess the effectiveness of a stakeholder engagement intervention in improving outcomes for communities affected by oil and gas extraction in Western Uganda	Data from other sources, such as documents, administrative data Community as a clustered unit for RCTs Equity; place of residence	No	Study design Quasi-experimental methods
Daidone <i>et al.</i> , (2017)	Publication details	Lesotho	Development sectors	To assess the impact of two	Poor households	No	Study design

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Peer-reviewed journal article Research report Country of first author Italy		Social protection Themes Human development and gender Social development and protection Urban and rural development	food security programmes in Lesotho	Households with children aged 0–17 years Equity; SES		Quasi-experimental methods
Dar <i>et al.</i> (2020)	Publication details Research report Country of first author USA	Bangladesh	Development sectors Agriculture, fishing and forestry Environment Themes Environment and natural resource management	To implement a randomised evaluation of a water management technique intervention in Bangladesh	Not specified	Yes	Study design RCTs
Edjekumhene <i>et al.</i> (2019)	Publication details Research report	Ghana	Development sectors Energy and extractives Governance, civil	To implement a randomised evaluation of Ghana's Public Interest and	Not specified	Yes	Study design cRCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Country of first author Ghana		society, and democracy Themes Public sector management Social development and protection	Accountability Committee information dissemination and engagement efforts			
Gibbs <i>et al.</i> (2020)	Publication details Peer-reviewed journal article Country of first author UK	Afghanistan	Development sectors Social protection Gender and social exclusion Themes Human development and gender	To assess the impact of the Women for Women International economic and social empowerment programme in Afghanistan	Adults only Experience with war/conflict Women Economic and social vulnerability Equity; SES, Sex, disability, conflict-affected	Yes	Study design Mixed methods: RCTs; interviews
Graves <i>et al.</i> (2018)	Publication details	Uganda	Development sectors	To implement a randomised	Attending control or	No	Study design

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Peer-reviewed journal article Country of first author USA		Health Themes Human development and gender	evaluation of a Family Clinic Day intervention in Uganda	intervention facilities in active care Equity; age; HIV/AIDS		Mixed methods: cRCTs; interviews; focus groups
Harris-Fry <i>et al.</i> (2018)	Publication details Peer-reviewed journal article Country of first author UK	Nepal	Development sectors Health Gender and social exclusion Themes Human development and gender	To implement a randomised evaluation of three interventions related to maternal health and nutrition	Adults only (as specified in the study) Pregnant women Equity; sex	Yes	Study design cRCTs
Hirvonen <i>et al.</i> (2017)	Publication details Peer-reviewed journal article	Ethiopia	Development sectors Health Themes	To assess the impact of caregivers' nutrition knowledge and its complementari	Not specified	No	Study design Quasi RCTs

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Country of first author Ethiopia		Human development and gender Urban and rural development	ty with market access and to test whether the effect of nutrition knowledge on children's dietary diversity depends on market access			
Källander <i>et al.</i> (2021)	Publication details Preprint, not peer-reviewed Country of first author UK, Sweden, USA	Uganda	Development sectors Health Themes Human development and gender	To present the results from a cRCT in Uganda which assessed the impact of appropriate treatment of children at cluster level of the two separate inSCALE interventions,	Adults only (as specified in the study)	Yes	Study design cRCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
				mHealth and VHCs, compared to control			
Khwaja <i>et al.</i> (2020)	Publication details Programme document Country of first author USA	Pakistan	Development sectors Public administration Themes Economic policy	To assess the impact of three types of tax policy interventions in Pakistan	Equity; sex, age	Yes	Study design Mixed methods: cRCTs; interviews and focus groups
Li and Liu (2020)	Publication details Research report Country of first author USA	China	Development sectors Transportation Themes Urban and rural development	To assess the impact of the rapid expansion of the subway system in Beijing, China	Data from other sources, such as documents, administrative data	Yes	Study design Quasi RCTs
McKenzie (2017)	Publication details Working paper	Nigeria	Development sectors Business, industry, trade and services	To implement a randomised evaluation of a national business	Adults only 18–40	Yes	Study design RCTs; quasi RCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Country of first author USA		Themes Private sector development	competition intervention in Nigeria			
Morten <i>et al.</i> (2020)	Publication details Programme document Country of first author USA	Tanzania	Development sectors Transportation Themes Urban and rural development Economic policy	To assess the impact of the Dar es Salaam Bus Rapid Transit System in Tanzania	Data from other sources, such as documents, administrative data Equity; place of residence	Yes	Study design Quasi RCTs
Mvukiyehe and van der Windt (2020)	Publication details Working paper Country of first author USA	Democratic Republic of the Congo	Development sectors Public administration Themes Human development and gender Public sector management Social development and protection	To implement a randomised evaluation of the long-term impact of a community-driven development programme	Not specified	No	Study design Mixed methods: RCTs; field experiment; interviews

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
Najy <i>et al.</i> (2018)	Publication details Peer-reviewed journal article Country of first author Pakistan	Benin	Development sectors Banking and financial services Themes Social development and protection Finance economic policy	To implement a randomised evaluation of three business registration interventions in Benin	Data from other sources, such as documents, administrative data	Yes	Study design Mixed methods: cRCTs; interviews; focus groups
Parker <i>et al.</i> (2019)	Publication details Programme document Country of first author USA	Uganda	Development sectors Energy and extractives Governance, civil society, and democracy Themes Public sector management	To assess the impact of stakeholder engagement interventions in Uganda	Equity; sex	Yes	Study design Mixed methods: RCTs; qualitative survey
Pellegrini (2018)	Publication details Programme document	Ecuador	Development sectors Water, sanitation, waste	To assess the impact of a quick and inexpensive	Not involved in another intervention/measurement	Yes	Study design RCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Country of first author Netherlands		Themes Human development and Gender Urban and rural development Environment and natural resource management	intervention that can improve water treatment, management and storage at the household level	population of at least 20 families/households Equity; place of residence		
Piper <i>et al.</i> (2018)	Publication details Peer-reviewed journal article Country of first author Kenya	Kenya	Development sectors Education Themes Human development and gender	To implement a randomised evaluation of three components of a literacy and numeracy intervention in the Kenyan context	Children and young people only (0–25 years old) Equity; age, education	No	Study design cRCTs
Rakotonari vo <i>et al.</i> (2017)	Publication details Peer-reviewed	Madagascar	Development sectors Environment Themes	To assess the validity of discrete choice experiments and the REDD+	Equity; place of residence	Yes	Study design Mixed methods: quasi RCTs; interviews

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	journal article Country of first author UK		Environment and natural resource management	deforestation project in Madagascar			
Robertson (2019)	Publication details Peer-reviewed journal article Country of first author USA	Cambodia	Development sectors Business, industry, trade and services Themes Private sector development	To assess the impact of a business compliance intervention in Cambodia	Data from other sources, such as documents, administrative data	Yes	Study design Quasi RCTs
Roth <i>et al.</i> (2017)	Publication details Working paper Country of first author Cambodia	Cambodia	Development sectors Banking and financial services Themes Finance	To assess the impact of microcredit access on paddy quantity and income in Cambodia	Not specified	No	Study design Quasi RCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
Saboya <i>et al.</i> (2018)	Publication details Programme document Country of first author Spain	Chad	Development sectors Humanitarian and emergencies Themes Human development and gender Social development and protection	To assess the impact of Chad's Protracted Relief and Recovery Operation	Children and young people only (children aged 6–23 months and their siblings aged 24–59 months)	Yes	Study design Mixed methods: quasi RCTs; interviews
Tsai <i>et al.</i> (2018)	Publication details Research report Country of first author USA	Philippines	Development sectors Social protection Governance, civil society, and democracy Themes Social development and protection	To implement a randomised evaluation of a civic leadership training programme	Adults only (as specified in the study)	Yes	Study design Mixed methods: cRCTs; interviews
Zegarra <i>et al.</i> (2017)	Publication details Working paper	Peru	Development sectors	To assess the impacts of a peer-to-peer training	Adults only Women	Yes	Study design Quasi RCTs

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Development sectors and themes	Study aims and objectives	Population characteristics	Theory/ theory of change	Research methodologies
	Country of first author Peru		Social protection Gender and social exclusion Themes Human development and gender Social development and protection	programme in Cañete Province, Peru	Women living in the province of Cañete Equity; sex; place of residence		

Table B: Key characteristics of systematic reviews

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
Akparibo <i>et al.</i> (2017)	Publication details Research report Country of first author UK	LMICs; countries facing humanitarian emergencies	Development Policy Sectors Health Humanitarian emergencies Themes Human development and gender Social development and protection	To investigate the relationship between recovery and relapse; and between relapse and default or return default/episodes of default in children aged 6–59 months affected by humanitarian emergencies	Children aged 6–59 months With health conditions (acute malnutrition)	No	Review approach Quantitative Methods of synthesis Narrative synthesis
Alampay <i>et al.</i> (2017)	Publication details Research report	LMICs	Development policy sectors Banking and financial services	To determine the impact of mobile financial services on the volume and frequency of remittances, the consumption habits of	Not specified	Yes	Review approach Quantitative Methods of synthesis

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	Country of first author Philippines		Information and communication Themes Finance	the poor and on livelihoods in terms of productivity and income			Meta-analysis
Ali <i>et al.</i> (2017)	Publication details Research report, user summaries Country of first author India	South Asia	Development policy sectors Governance, civil society, and democracy Themes Social development and protection	To explore different models of non-state justice systems in South Asia and different approaches for strengthening complementarity between state and non-state justice delivery. To assess the effects of these interventions	Not specified	No	Review approach Qualitative Methods of synthesis Thematic narrative analysis

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
Annamalai <i>et al.</i> (2017)	<p>Publication details Research report</p> <p>Country of first author India</p>	LMICs	<p>Development policy sectors Transportation Water, sanitation and waste</p> <p>Themes Urban and rural development</p>	To assess the effectiveness of interventions which seek to improve access to and quality of civic infrastructure and services	Not specified	Yes	<p>Review approach Mixed methods</p> <p>Methods of synthesis Narrative synthesis; descriptive numerical summary approach and cause and effect analysis</p>
Babu <i>et al.</i> (2017)	<p>Publication details Research report</p> <p>Country of first author</p>	South Asia	<p>Development policy sectors Social protection</p> <p>Themes Social development</p>	To assess the effects of various interventions and approaches used for enhancing poverty reduction and development benefits of within-country migration	Migrants	Yes	<p>Review approach Mixed methods</p> <p>Methods of synthesis Meta-analysis; narrative</p>

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	author India		and protection				synthesis; count of evidence
Blundo-Canto <i>et al.</i> (2018)	Publication details Peer-reviewed journal article Country of first author France	Global/worldwide	Development policy sectors Agriculture, fishing and forestry Environment Themes Environment and natural resource management	To analyse evidence of the livelihoods impacts of Payments for Environmental Services (PES)	Not specified	Yes	Review approach Scoping Methods of synthesis Reporting trends (descriptive)
Catalano <i>et al.</i> (2019)	Publication details Research report, peer-reviewed	LMICs	Development policy sectors Education Social protection	To explore how positive youth development approaches have been implemented in LMICs and to assess the effectiveness of these approaches	Children and young people only (0–25 years old)	Yes	Review approach Quantitative Methods of synthesis

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	journal article Country of first author USA		Themes Human development and gender Social development and protection				Descriptive approach
De Buck <i>et al.</i> (2017)	Publication details Research report Country of first author Belgium	LMICs	Development policy sectors Health Water, sanitation, waste Themes Human development and gender Urban and	To assess the effectiveness of, and influencing factors for, different approaches for promoting handwashing and sanitation behaviour change in communities in LMICs	Not specified	Yes	Review approach Mixed methods Methods of synthesis Meta-analysis; best fit framework synthesis

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
			rural development				
Garn <i>et al.</i> (2017)	Publication details Peer-reviewed journal article Country of first author USA	Global/worldwide	Development policy sectors Water, sanitation, waste Themes Urban and rural development	To characterise the impacts of various sanitation interventions on latrine coverage and also on latrine use. To explore how various structural and design characteristics of sanitation (e.g. smell, presence of a door, etc.) were reported to be associated with use of latrines	Not specified	No	Review approach Mixed methods Methods of synthesis Meta-analysis; narrative synthesis
Ghose <i>et al.</i> (2017)	Publication details Research report Country of first	LMICs	Development policy sectors Energy and extractives Themes	To assess the efficacy of natural resource funds as an intervention to manage revenues from mineral resources in LMICs experiencing	Experiencing political instability	Yes	Review approach Mixed methods Methods of synthesis

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	author India		Finance Environment and natural resource management	politically fragile circumstances			Narrative synthesis
Hossain <i>et al.</i> (2017b)	Publication details Research report Country of first author Bangladesh	LMICs; South Asia; East Asia	Development policy sectors Humanitarian emergencies Themes Urban and rural development	To explore and understand the complex relationships between urbanisation, natural disasters and vulnerability. To identify effective efforts and processes that address this complexity and contribute to mitigating the risks of natural disaster. To effectively disseminate lessons learned and current best practices found in the evidence base	Not specified	Yes	Review approach Mixed methods Methods of synthesis Framework synthesis

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
Hossain <i>et al.</i> (2017a)	Publication details Research report Country of first author Bangladesh	LMICs	Development policy sectors Social protection Themes Social development and protection	To assess the impact of approaches for addressing insecurity or violence arising from urbanisation	Not specified	Yes	Review approach Quantitative Methods of synthesis Narrative synthesis
Ilavarasan <i>et al.</i> (2017b)	Publication details Research report Country of first author India	LMICs	Development policy sectors Business, industry, trade and service Information and communication	To examine whether access to business-relevant information through networked devices enhances the internal efficiency and business growth of urban micro-, small and medium enterprises in LMICs	Only urban areas	Yes	Review approach Quantitative Methods of synthesis Meta-analysis; narrative synthesis

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
			Themes Finance				
Ilavarasan <i>et al.</i> (2017a)	Publication details Research report Country of first author India	South Asia	Development policy sectors Business, industry, trade and services Themes Private sector development	To contextualise the findings of the evidence summary emerging out of 11 systematic reviews on employment outcomes of skills training in South Asian countries. To assess what types of skills training have shown most impact on employment outcomes in LMICs, particularly in South Asia	Not specified	No	Review approach Systematic review of reviews Methods of synthesis Narrative synthesis Contextual analysis
Kumar <i>et al.</i> (2018)	Publication details Research report	LMICs; South Asia	Development policy sectors Agriculture, fishing and forestry	To assess the effectiveness of market-led development approaches among the rural and semi-urban population in LMICs	Rural and semi-urban population Rural population	Yes	Review approach Mixed methods Methods of synthesis

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	Country of first author USA		Business, industry, trade and services Themes Private sector development Social development and protection Urban and rural development	and to assess the factors that determine the success of different market-led development approaches in subsistence and migrant-driven rural economies			Meta-analysis; narrative synthesis
Langer <i>et al.</i> (2018)	Publication details Research report Country of first	LMICs	Development policy sectors Business, industry, trade and services	To produce an interactive evidence map of research evaluating interventions aiming to overcome barriers to women's economic	Women	Yes	Review approach Quantitative Methods of synthesis

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	author South Africa		Gender and social exclusion Themes Finance Human development and gender	empowerment in LMICs. To synthesise evidence on the effects of interventions supporting women's participation in wage labour in higher-growth and/or male-dominated sectors in LMICs. To identify design features that influence the effects of interventions aiming to overcome barriers to women's economic empowerment in LMICs			Meta-analysis; narrative synthesis
Maynard <i>et al.</i> (2017)	Publication details Research report	LMICs	Development policy sectors Social protection Themes	To investigate the process of implementing humanitarian interventions supporting shelter self-	Populations affected by humanitarian crises	Yes	Review approach Mixed methods Methods of synthesis

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	Country of first author UK		Social development and protection	recovery, and to assess their effectiveness			Narrative synthesis
Menon <i>et al.</i> (2018)	Publication details Research report Country of first author India	LMICs	Development policy sectors Health Themes Human development and gender	To identify, critically appraise and provide an overview of review-level evidence on the effectiveness of nutritional interventions delivered in LMICs targeted at the World Health Assembly specified outcomes	Women	Yes	Review approach Quantitative Methods of synthesis Narrative synthesis
Nair <i>et al.</i> (2017b)	Publication details Research report Country of first author	LMICs	Development policy sectors Gender and social exclusion	To identify the principles, components and theories of change of interventions to enhance the gender responsiveness of policing. To synthesise	Age 15 and above Women (the participants included men and women aged 15 or above and	Yes	Review approach Mixed methods Methods of synthesis

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	author India		Themes Human development and gender	evaluations of these interventions and assess reception by the target and participant groups	transgender population, in relation to whom Gender Responsive Policing interventions were implemented)		Meta-analysis; framework synthesis
Nair <i>et al.</i> (2017c)	Publication details Research report Country of first author India	LMICs	Development policy sectors Education Banking and financial services Themes Finance	To assess the effectiveness of public works programmes in stimulating local economic transformation in LMICs	Not specified	Yes	Review approach Contextualisation analysis Methods of synthesis Contextual analysis
Nair <i>et al.</i> (2017a)	Publication details Research report	South Asia; global/worldwide	Development policy sectors Health Information	To contextualise findings from systematic reviews and to apply them to Nepal and Bangladesh	Adults only Pregnant women	Yes	Review approach Contextualisation analysis

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	Country of first author India		and communication Themes Human development and gender				Methods of synthesis Contextual analysis
Nidhi <i>et al.</i> (2017)	Publication details Research report Country of first author India	Global/worldwide	Development policy sectors Humanitarian emergencies Environment Themes Urban and rural development	To summarise review-level evidence on the effectiveness of disaster management approaches in LMICs, and to contextualise the evidence to provide policymakers with a reliable basis for informed decision-making regarding the applicability and transferability of different disaster	Not specified	Yes	Review approach Systematic review of reviews Methods of synthesis Narrative synthesis

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
				management interventions to South Asian settings, with particular emphasis on Bangladesh			
Obuku <i>et al.</i> (2017)	Publication details Research report Country of first author Uganda	Fragile and post-conflict states	Development policy sectors Health Themes Human development and gender	To assess the effectiveness of different approaches to engaging with non-state providers in improving the delivery of primary healthcare in fragile, conflict or post-conflict settings	Not specified	No	Review approach Quantitative
Patel <i>et al.</i> (2017)	Publication details Research report Country of first	LMICs; countries facing humanitarian emergencies	Development policy sectors Humanitarian emergencies Themes	To present the evidence on practices to identify and prioritise vulnerable populations affected by urban humanitarian emergencies	Vulnerability Only urban areas	No	Review approach Mixed methods Methods of synthesis Thematic synthesis

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	author USA		Urban and rural development				
Pilkington <i>et al.</i> (2017)	Publication details Research report Country of first author UK	LMICs	Development policy sectors Health Themes Human development and gender	To assess the effectiveness of community engagement/participation approaches for delivering better health outcomes, improving service delivery and sustaining benefits	Not specified	Yes	Review approach Quantitative Methods of synthesis Narrative synthesis
Stone <i>et al.</i> (2020)	Publication details Research report Country of first author USA	Latin America and the Caribbean	Development policy sectors Education Themes Human development and gender	To assess the impact of reading programmes, practices, policies, and products aimed at improving the reading skills of children from birth through Grade 3 on reading outcomes in the Latin America and the Caribbean region	Children and young people only (0–25 years old)	Yes	Review approach Mixed methods Methods of synthesis Meta-analysis; thematic synthesis

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
Waddington <i>et al.</i> (2019)	<p>Publication details Research report</p> <p>Country of first author UK</p>	LMICs	<p>Development policy sectors Public administration Governance, civil society, and democracy</p> <p>Themes Public sector management</p>	To identify and evaluate the effects of interventions that aim to strengthen participatory, inclusive, transparent or accountable mechanisms on social and economic wellbeing of participants and participatory, inclusive, transparent or accountable processes	Not specified	Yes	<p>Review approach Mixed methods</p> <p>Methods of synthesis Meta-analysis; framework synthesis</p>
Williamson <i>et al.</i> (2017)	<p>Publication details Research report</p> <p>Country of first author UK</p>	LMICs	<p>Development policy sectors Social protection</p> <p>Themes Social development</p>	To assess the impact of protection interventions on unaccompanied and separated children, during the period of separation, in	<p>Children under 18 years old</p> <p>Separated children</p>	Yes	<p>Review approach Mixed methods</p> <p>Methods of synthesis Narrative synthesis</p>

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
			and protection	humanitarian crises in LMICs			
Wolf <i>et al.</i> (2018)	Publication details Peer-reviewed journal article Country of first author Switzerland	LMICs	Development policy sectors Water, sanitation, waste Themes Urban and rural development	To provide an updated assessment of the impact of unsafe water, sanitation and hygiene on childhood diarrhoeal disease	Children under five years old	Yes	Review approach Quantitative Methods of synthesis Meta-analysis
Yount <i>et al.</i> (2017)	Publication details Peer-reviewed journal article	LMICs	Development policy sectors Health Themes	To synthesise evidence from systematic reviews on the impact of interventions to prevent violence and victimisation against adolescent girls and	Aged 10–24 years old Adolescent girls and young women	Yes	Review approach Quantitative; systematic review of reviews

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Short title	Publication	Country	Policy sectors and themes	Study aims/objectives	Population characteristics	Theory / theory of change	Methodology
	Country of first author USA		Human development and gender	young women aged 10–24 years in LMICs			Methods of synthesis Narrative synthesis
Zwi <i>et al.</i> (2018)	Publication details Research report Country of first author Australia	LMICs	Development policy sectors Humanitarian emergencies Themes Social development and protection	To assess the impact of community-based disaster risk management initiatives on the social and economic costs of disasters	Not specified	Yes	Review approach Mixed methods Methods of synthesis Realist mapping and review

ANNEX E: Social values in health and care guidance

From Gough *et al.* (2014).

SV1: Utility and efficiency (effectiveness and cost effectiveness)

Issue 1: *Whether a consequentialist cost effectiveness approach is warranted.*

Issue 2: *Whether the social value of rights to health and welfare outcomes should be adopted.*

Issue 3: *Whether there should be a broader conception of the interventions in cost effectiveness evaluations.*

Issue 4: *Whether there should be a broader conception of outcomes in cost effectiveness interventions.*

SV2: Justice and equity

Issue 5: *The extent that all possible weightings by group, situation and outcome (as, for example, in the lists of ideas above) should be specified in guidance in terms of:*

(i) identifying factors that could be taken into account in particular types of circumstances;

(ii) proposing specific balances between utility and such equity weightings;

(iii) specification of how to manage tensions between different social values weightings in the same case;

(iv) whether any weightings should be applied specifically to reduce inequalities in society.

Issue 6: *Whether all guidance production should include an assessment of both explicit and possible inadvertent social value weightings with, for example, a standard instrument or checklist to help make such assessments.*

Issue 7: *Whether to be more specific about how community and individual and group needs and guidance relate and whether community needs should also relate to global need.*

Issue 8: *Considering the social values of how to balance competing social values.*

Issue 9: *Considering developing social values for emergency and other special situations.*

Issue 10: *Considering the balance of social value and thus weighting for innovative services.*

Issue 11: *Whether human values such as dignity, compassion, commitment and human relationships should become part of the guidance process.*

SV3: Autonomy

Issue 12: *The extent of explicit advice on the social values underpinning the tension between guidance and professional and client decision-making (such as the goals of the guidance) and how these might be resolved.*

SV4: Solidarity

Issue 13: *Whether there should be explicit advice on the social values underpinning the tension between public outcomes and individual benefit and individual autonomy.*

Issue 14: *Whether to explicitly incorporate the value of solidarity in terms of social inequality and social values.*

SV5: Participation

Issue 15: *Whether to be more explicit and evidence-informed about participation.*

SV6: Sustainability

Issue 16: *Considering sustainability issues in terms of the ability to provide ongoing guidance in its financial, political and broader societal contexts.*

Guidance producers considering what processes they are able to manage, not just now but in an ongoing sustainable way.

Issue 17: *Considering sustainability issues in terms of the environmental and other impacts of the implementation of its guidance.*

Issue 18: *Considering sustainability issues in terms of the wellbeing of and costs for future generations.*

SV7: Transparency and accountability

Issue 19: *Whether accountability and transparency is sufficiently and correctly specified in guidance processes.*

SV8: Appropriate methods of guidance development

Issue 20: *Whether there should be formal processes for examining the social values being applied in the whole guidance development process and specific stages of: topic identification and clarification; evidence identification and analysis; calculation of metrics of cost effectiveness; guidance decisions; and implementation.*

Issue 21: *Considering the impact of lack of research data on, for example, harms, distributional effects, and implementation.*

Issue 22: *Considering the overall fitness of purpose and timeliness of guidance.*

ANNEX F: Framing of systematic reviews to make evidence claims

Review type	Who framed the review		Process of framing review		Framing and social values apparent in review
	Review team	Team plus	Literature	Stakeholders	
Meta-analysis					
Alampay <i>et al.</i> (2017)	x				<ul style="list-style-type: none"> Outcomes of financial and social inclusion (Alampay <i>et al.</i>, 2017)
Stone <i>et al.</i> (2020)	x				<ul style="list-style-type: none"> Education (Stone <i>et al.</i>, 2020)
Wolf <i>et al.</i> (2018)	x				<p>Indicators matching SDG targets:</p> <ul style="list-style-type: none"> Solidarity and rights (SDG 6 water) Wolf <i>et al.</i> (2018)
	3/3	0/3	0/3	0/3	
Qualitative synthesis					
Ali <i>et al.</i> (2017)	x				<p>In the question: Efficiency and accessibility of informal justice systems (Ali <i>et al.</i>, 2017)</p>
	1/1	0/1	0/1	0/1	
Framework synthesis					
Hossain <i>et al.</i> (2017b)	x		x		<p>Theory of change with values:</p> <ul style="list-style-type: none"> Equity, community engagement and resilience (Hossain <i>et al.</i>, 2017b)
Williamson <i>et al.</i> (2017)	x		x		<p>Frameworks for categorising interventions:</p> <ul style="list-style-type: none"> Child rights, ecological systems theory, vulnerability and resilience (Williamson <i>et al.</i>, 2017)

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Review type	Who framed the review		Process of framing review		Framing and social values apparent in review
	Review team	Team plus	Literature	Stakeholders	
	2/2	0/2	2/2	0/2	
Meta-analysis plus					
Garn <i>et al.</i> (2017)	x				Outcomes <ul style="list-style-type: none"> • Utility, sustainability (Garn, 2017)
Waddington <i>et al.</i> (2019)	x				<ul style="list-style-type: none"> • Participation, solidarity, equity, accountability, social and economic development, sustainability (Waddington, 2019)
Nair <i>et al.</i> (2017a)		x			Ecological model <ul style="list-style-type: none"> • Gender and justice (Nair <i>et al.</i>, 2017a)
De Buck <i>et al.</i> (2017)		x		x	Theory of change with values: <ul style="list-style-type: none"> • Solidarity and rights (SDG 6 water); cultural sensitivity (De Buck <i>et al.</i>, 2017)
Langer <i>et al.</i> (2018)		x	x	x	Design features of interventions <ul style="list-style-type: none"> • Gender equity, participation in employment, social capital, economic empowerment (Langer <i>et al.</i>, 2018)
	2/5	3/5	1/5	2/5	
Narrative review					
Akparibo <i>et al.</i> (2017)	x				<ul style="list-style-type: none"> • Compliance with programme (Akparibo <i>et al.</i>, 2017)
Patel <i>et al.</i> (2017)	x				<ul style="list-style-type: none"> • Equity (Patel <i>et al.</i>, 2017)
Blundo-Canto <i>et al.</i> (2018)	x		x		<ul style="list-style-type: none"> • Sustainability/solidarity, individualism, heritage and self-determination, equity (Blundo-Canto <i>et al.</i>, 2018)

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Review type	Who framed the review		Process of framing review		Framing and social values apparent in review
	Review team	Team plus	Literature	Stakeholders	
Kumar <i>et al.</i> (2018)	x				<ul style="list-style-type: none"> Economic and social outcomes (Kumar <i>et al.</i>, 2018)
Hossain <i>et al.</i> (2017a)	x		x		<p>Ecological model</p> <ul style="list-style-type: none"> Justice, equity, safety/protection (Hossain, 2017a)
Ghose <i>et al.</i> (2017)		x	x	x	<ul style="list-style-type: none"> Transparency and accountability, equality and maintaining peace, macroeconomic stability (Ghose, 2017) <p>Contextualisation</p> <ul style="list-style-type: none"> Gender inequality, community mobilisation (Hossain, 2018)
Catalano Richard <i>et al.</i> (2019)		x	x	x	<p>Logic model</p> <ul style="list-style-type: none"> Solidarity, individualism (Catalano Richard <i>et al.</i>, 2019)
Maynard <i>et al.</i> (2017)		x	x	x	<p>Theory of change:</p> <ul style="list-style-type: none"> Efficiency, effectiveness (Maynard <i>et al.</i>, 2017) <p>In findings</p> <ul style="list-style-type: none"> Household dignity, self-reliance, safety, assets/debts, livelihoods, health (Maynard <i>et al.</i>, 2017)
	5/8	3/8	6/8	3/8	
Realist review					
Zwi <i>et al.</i> 2018		x	x	x	<p>Proposed mechanisms for change:</p> <ul style="list-style-type: none"> Integrated knowledges, community empowerment, actioned agency

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Review type	Who framed the review		Process of framing review		Framing and social values apparent in review
	Review team	Team plus	Literature	Stakeholders	
					In findings: <ul style="list-style-type: none"> Resilient livelihoods, gender and social equity, safety, security and protection. Technological innovation and communication (Zwi <i>et al.</i>, 2018)
	0/1	1/1	1/1	1/1	
Review of reviews					
Babu <i>et al.</i> (2017)	x				<ul style="list-style-type: none"> Equity, solidarity and individualism (Babu <i>et al.</i>, 2017)
Ilavarasan <i>et al.</i> (2017b)	x				<ul style="list-style-type: none"> Participation and equity in employment (Ilavarasan <i>et al.</i>, 2017b)
Ilavarasan <i>et al.</i> (2017a)	x				<ul style="list-style-type: none"> Business communication, efficiency and growth; context, gender and culture (Ilavarasan, 2017a)
Srivastava <i>et al.</i> (2017)	x				In contextualisation: <ul style="list-style-type: none"> Efficiency, participation, cultural norms and traditional beliefs, inequity, vulnerability (Srivastava <i>et al.</i>, 2017)
Obuku <i>et al.</i> (2017)	x				In findings and recommendations Participation, solidarity, efficiency (Obuku <i>et al.</i> , 2017)
Pilkington <i>et al.</i> (2017)	x				<ul style="list-style-type: none"> Participation, norms, power, sustainability (Pilkington <i>et al.</i>, 2017)
Yount <i>et al.</i> (2017)	x		x		Ecological model

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Review type	Who framed the review		Process of framing review		Framing and social values apparent in review
	Review team	Team plus	Literature	Stakeholders	
Annamalai <i>et al.</i> (2017)		x	x	x	<ul style="list-style-type: none"> • Solidarity (quality of public infrastructure), equity (of access) Annamalai <i>et al.</i> (2017)
Menon <i>et al.</i> (2018)		x	x	x	<p>Focus/ question</p> <ul style="list-style-type: none"> • Solidarity: World Health Assembly undernutrition targets set for 2025 (Menon <i>et al.</i>, 2018)
Nair <i>et al.</i> (2017b)		x	x	x	<ul style="list-style-type: none"> • Effectiveness, community participation (Nair <i>et al.</i>, 2017b) <p>Contextualisation</p> <ul style="list-style-type: none"> • Ecological model
Nair <i>et al.</i> (2017c)		x	x	x	<p>Contextualisation</p> <ul style="list-style-type: none"> • Ecological model
	7/11	4/11	5/11	4/11	
Grand total	20/31	11/31	15/31	10/31	

Annex G: Truth table of attributes for developing evidence claims

	Systematically debating: focus, scope, or backdrop		Social values integrated into evidence			Scope of claims	Social impact
Study	With Stakeholders	Formal consensus	Intervention/ theory of change	Outcomes of interest	Other mentions	Certainty and reach	
Informing policy or guidance: instrumental impact							
De Buck <i>et al.</i> (2017) Meta-analysis plus	Advisory group, but not described. Funded by Water Supply and Collaborative Council	SDG 6: water	Solidarity and rights to water; cultural sensitivity			CASP GRADE Identified key facilitators	WHO methods guide for evidence synthesis for health policy and systems
Garn <i>et al.</i> (2017) Meta-analysis plus	WHO-funded		Sanitation privacy and cleanliness	Utility, sustainability		Liverpool Quality Appraisal Tool GRADE = low Generalisable Forest plot, with heterogeneity	WHO guidelines
Langer <i>et al.</i> (2018)	Advisory group: policymakers		Gender equity, participation			Cochrane ROB GRADE	World Bank nutrition portfolio,

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Meta-analysis plus	and academics		in employment, social capital, economic empowerment			Identified seven active intervention components	World Bank (2021)
Maynard <i>et al.</i> (2017) Narrative review	Oxfam			Household dignity, self-reliance, safety, assets/debts, livelihoods, health		Quality assessment tool Consistent effects on: households' dignity, self-reliance, safety and security. Inconsistent: household incomes/ livelihoods, assets/ debts, physical and mental health, and knowledge of safer construction techniques	VEHA Guidance , Promoting Safer Building Working Group, ShelterCluster.org , ReliefWeb , Humanitarian Library , Issue Lab , ALNAP

<p>Nair <i>et al.</i> (2017b) Gender-responsive policing Meta-analysis plus</p>	<p>Members of police service</p>		<p>Corruption, gender inequality, rights, empowerment, solidarity (collectivisation)</p>	<p>Gender, justice</p>		<p>CASP, MMAT, Newcastle-Ottawa, Scope = LMICs</p>	<p>GBV AoR Help Desk</p>
<p>Waddington <i>et al.</i> (2019) Meta-analysis plus</p>	<p>Advisory group of academics and policymakers with specific expertise in governance</p>	<p>SDG 16: Peaceful and inclusive societies, access to justice and effective, accountable and inclusive institutions</p>		<p>Participation, solidarity, equity, accountability, social and economic development, sustainability</p>		<p>ROB drew from: Hombrados and Waddington (2012) and bias domains and extensions to Cochrane's ROBINS-I tool and RoB2.0 (Sterne <i>et al.</i>, 2016; Higgins <i>et al.</i>, 2016). Not generalisable because of the small sample of studies</p>	<p>WHO Guide on citizen engagement with EIPM</p>

Wolf <i>et al.</i> (2018) Meta-analysis	First author at WHO	SDG 6: water	Access to drinking water, sanitation and hygiene	Drinking water, sanitation and hygiene		Newcastle-Ottawa Scale, Forest plot: generalisable	WHO guidance
Zwi <i>et al.</i> (2018) Realist review	Reference group (p. 21)		Integrated knowledges, community empowerment, actioned agency		Resilient livelihoods, gender and social equity, safety, security and protection, technological innovation and communication	Quality assessment tool Realist review offers theoretical generalisability	Prevention web , Investing in Rural People Training Manual
Informing understanding: conceptual impact							
Blundo-Canto <i>et al.</i> (2018) Narrative review	Lead author affiliated with CIRAD, a French agency that mobilises science, innovation		Environmental/ecosystem services		Sustainability/solidarity, individualism, heritage and self-	Quality assessment tool	Food and Agriculture Organization of the United Nations

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

	and training in order to achieve the SDGs				determination, equity		
Catalano Richard <i>et al.</i> (2019) Narrative review	Authors include NGOs who compile and use evidence		Solidarity, individualism			Quality assessment tool	WHO Knowledge Summary
Contributing to knowledge repositories							
Akparibo (2017) Narrative review	Oxfam	None	None	None		CASP	ALNAP , ReliefWeb , 3ie evidence hub , IssueLab
Patel <i>et al.</i> (2017) Narrative review	Oxfam	None	None	None	Equity	Quality assessment tool Re	Save the Children , ALNAP , Issue Lab
Williamson <i>et al.</i> (2017) Framework synthesis	Advisory group of the Inter-Agency Working Group on Unaccompanied and Separated Children		Child rights, ecological systems theory, vulnerability and resilience	Pro-social behaviour		Quality assessment tool CASP Mentions over-emphasis on conflict in Africa	Issue lab. , ReliefWeb , Save the Children , Child Protection Hub , Better Care Network , ALNAP etc.

Stone <i>et al.</i> (2020) Meta-analysis	None	Consensus conference: Education for All (1990)	Educational practices; nutrition programmes	Literacy		CASP, Waddington and Hombrados (2012) tool Theory of change: combinations and conditions leading to +/- effects	UNESCO's IIEP Learning Portal , AIR, NIH, IDEAS
Kumar <i>et al.</i> (2018) Narrative review	None	None	None	Economic and social outcomes		Cochrane ROB Transferability to Nepal	CABI
No apparent impact							
Nair <i>et al.</i> (2017c) Public works Meta-analysis	Advisory group includes a national bank and a civil society organisation		Social cohesion, corruption, participation, empowerment, community involvement	Poverty, employment, debt, welfare, social unrest, participation, water and soil		CASP, MMAT. EPOC Transferability to Nepal	None found
Ali <i>et al.</i> (2017) Qualitative synthesis	Advisory group: members of justice		Access to justice	Access, efficiency, gender, equity, accountability,		Quality assessment No theoretical or statistical	None found

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

	system, and academics			restorative justice, reduced crime, human rights		judgements of transferability. Transferability to Nepal	
Alampay <i>et al.</i> (2017) Meta-analysis			Financial/ social inclusion	Financial/ social inclusion	Utility, rights	Quality assessment: Waddington and Hombrados (2012) tool. Forest plot: (in)consistency 'generally' promising	None found
Hossain <i>et al.</i> (2017b) Framework synthesis			Vulnerability, community engagement and resilience (p. 18)			Quality assessment tool Medium- to high-quality evidence No certain effects reported Transferability to Nepal and Bangladesh	None found

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Hossain <i>et al.</i> (2018) Narrative review			Equity, community engagement and resilience		Ecological model: justice, equity, safety/ protection	Quality assessment tool Transferability to South Asia	None found
Ghose <i>et al.</i> (2018) Narrative review	Retired senior officials from the Government of India				Transparency and accountability, equality and maintaining peace, macroeconomic stability; contextualisation: gender inequality, community mobilisation	Quality assessment tool Attempted to transfer high-income country evidence to Myanmar and Afghanistan	None found
Pilkington <i>et al.</i> (2017)	Advisory group includes ex-				Participation, norms, power,	AMSTAR Transferability to Nepal	None found

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

Review of reviews	government minister				sustainability		
Menon <i>et al.</i> (2018) Review of reviews		World Health Assembly Global Nutrition Targets 2025				AMSTAR Generalisable to LMICs	None found
Obuku <i>et al.</i> (2017) Review of reviews				Community empowerment	Participation, solidarity, efficiency	Effective Public Health Practice Project quality assessment tool	None found
Yount <i>et al.</i> (2017) Review of reviews			Women's empowerment			AMSTAR	None found
Ilavarasan <i>et al.</i> (2017a) Review of reviews					Participation and equity in employment	AMSTAR Transferability to Nepal, India and LMICs	None found
Ilavarasan (2017b) Review of reviews					Business communication, efficiency	Cochrane ROB	None found

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

					and growth; context, gender and culture		
Srivastava <i>et al.</i> (2017) Review of reviews					Contextualisation: efficiency, participation, cultural norms and traditional beliefs, inequity, vulnerability	AMSTAR	None found
Babu <i>et al.</i> (2017) Review of reviews					Equity, solidarity and individualism	Quality assurance South Asia	None found
Annamalai <i>et al.</i> (2017) Review of reviews					Solidarity (quality of public infrastructure),	Quality assessment tool	None found

CEDIL syntheses working paper 10: Making and justifying evidence claims: Evidence synthesis of impact evaluations and systematic reviews in international development

					equity (of access)		
Nair <i>et al.</i> (2017a) Behaviour change Review of reviews					Effectiveness, community participation	Quality assessment tool Transferability to Nepal and Bangladesh	None found

References

- Ahmed, A., Hoddinott, J. & Roy, S. 2019. Food Transfers, Cash Transfers, Behavior Change Communication and Child Nutrition Evidence from Bangladesh, IFPRI Discussion Paper 1868. Washington, DC: International Food Policy Research Institute (IFPRI). <https://doi.org/10.2499/p15738coll2.133420>,
- Aker, J., Collier, P. & Vicente, P. C. 2017. Is Information Power? Using Mobile Phones and Free Newspapers During an Election in Mozambique. *Review of Economics and Statistics*, 99, 185-200.
- Aker, J. C. & Ksoll, C. 2019. Call me educated: Evidence from a mobile phone experiment in Niger. *Economics of Education Review*, 72, 239-257.
- Akparibo, R., Booth, A. & Lee, A. 2017. Recovery, Relapse, and Episodes of Default in the Management of Acute Malnutrition in Children in Humanitarian Emergencies: A systematic review. Humanitarian Evidence Programme. Oxford: Oxfam GB.
- Alampay, E. A., Moshi, G. C., Ghosh, I., Peralta, M. L. C. & Harshanti, J. 2017. The impact of mobile financial services in low- and lower middle-income countries. **International Development Research Centre, Ottawa, Canada, and the Department for International Development, UK.**
- Ali, F., Mathew, S. K., Gopaldaswamy, A. K. & Babu, M. S. 2017. Systematic review of different models and approaches of non-state justice systems in South Asia and its complementarity with the state justice delivery systems. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London..
- Amirapu, A., Asadullah, N. M. & Wahhaj, Z. 2020. Can Child Marriage Law Change Attitudes and Behaviour? Experimental Evidence from an Information Intervention in Bangladesh. School of Economics Discussion Papers, University of Kent, UK.
- Angeles, G., Handa, S., Spektor, G., Chakrabarti, A. & Otchere, F. 2019. Impacts of Zimbabwe's Harmonised Social Cash Transfer Programme. 3ie Grantee Final Report. New Delhi: International Initiative for Impact Evaluation (3ie).
- Annamalai, T. R., Devkar, G. A., Delhi, V. S. K., Ramanarayanan, V. & Goswami, S. 2017. How effective are interventions which seek to improve access and quality of civic infrastructure and amenities? What are the key characteristics of successful interventions in urban areas – A Rapid Evidence Summary. London UK. EPPI Centre, Social Science Research Unit, Institute of Education, University of London
- Armand, A., Costa, A. I., Coutts, A., Vicente, P. & Vilela, I. 2019. Using Information to Break the Political Resource Curse in Natural Gas Management in Mozambique.., 3ie Impact Evaluation Report 93. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: <https://doi.org/10.23846/TW8IE93>

- Asfaw, S., Carraro, A., Davis, B., Handa, S. & Seidenfeld, D. 2017. Cash transfer programmes, weather shocks and household welfare: evidence from a randomised experiment in Zambia. *Journal of Development Effectiveness*, 9, 419-442.
- Asunka, J., Brierley, S., Golden, M., Kramon, E. & Ofosu, G. 2019. Electoral Fraud or Violence – The Effect of Observers on Party Manipulation Strategies. *Other*, 49, 129-151.
- Avdeenko, A. & Frölich, M. 2019. Impacts of Increasing Community Resilience through Humanitarian Aid in Pakistan, 3ie Impact Evaluation Report 100. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: <https://doi.org/10.23846/TW6IE100>
- Babu, M. S., Gopalaswamy, A. K., Baskar, V. & Dash, U. 2017. Effects of interventions and approaches for enhancing poverty reduction and development benefits of 'within country migration' in South Asia. A systematic review. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I. & Sulaiman, M. 2017. Labor Markets and Poverty in Village Economies. *Quarterly Journal of Economics*, 132, 811-870.
- Banerjee, A., Barnhardt, S. & Duflo, E. 2018. Can Iron-Fortified Salt Control Anemia? Evidence from Two Experiments in Rural Bihar. *Journal of Development Economics*, 001-072.
- Bangpan, M., Draper, A., Minchenko, D., Stansfield, C., Dickson, K., Tripney, J. & Oliver, S. 2022. Making and Justifying evidence claims: A protocol of evidence synthesis of impact evaluations and systematic reviews in international development. CEDIL methods working paper. London: Centre of Excellence for Development Impact and Learning.
- Barnett, C., Masset, E., Dogbe, T., Jupp, D., Korboe, D., Acharya, A., Nelson, K., Eager, R. & Hilton, T. 2018. Impact Evaluation of the SADA Millennium Villages Project in Northern Ghana: Endline Summary Report, ' Brighton: Itad in association with IDS, LSHTM and PDA Ghana
- Belissa, T., Bulte, E., Cecchi, F., Gangopadhyay, S. & Lensink, R. 2019. Liquidity Constraints, Informal Institutions, and the Adoption of Weather Insurance: A Randomized Controlled Trial in Ethiopia. *Journal of Development Economics*, 140, 269-278.
- Béné, C., Riba, A. & Wilson, D. 2020. Impacts of resilience interventions – Evidence from a quasi-experimental assessment in Niger. *International Journal of Disaster Risk Reduction*, 43.
- Berg, E., Ghatak, M., Manjula, R., Rajasekhar, D. & Roy, S. 2019. Motivating Knowledge Agents: Can Incentive Pay Overcome Social Distance? *The Economic Journal*, Volume 129, Issue 617, January 2019, Pages 110–142, <https://doi.org/10.1111/eoj.12544>

- Berhane, G., Hoddinott, J., Kumar, N. & Margolies, A. 2017. The Productive Safety Net Programme in Ethiopia Impacts on Children's Schooling, Labour and Nutritional Status, 3ie Impact Evaluation Report 55. New Delhi: International Initiative for Impact Evaluation (3ie),
- Bero, L., Chartres, N., Diong, J., Fabbri, A., Gherzi, D., Lam, J., Lau, A., McDonald, S., Mintzes, B. & Sutton, P. 2018. The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures. *Systematic Reviews*, 7, 1-11.
- Bett, E., Mugwe, J., Nyalugwe, N., Haraman, E., Williams, F., Tambo, J., Wood, A. & Bundi, M. 2018. Impact of plant clinics on disease and pest management, tomato productivity and profitability in Malawi, CABI Working Paper 11, 30pp.
- Blattman, C., Green, D., Ortega, D. & Tobón, S. 2018. Hotspot Interventions at Scale: The Effects of Policing and City Services on Crime in Bogotá, Colombia, . Hotspot Interventions at Scale: The Effects of Policing and City Services on Crime in Bogotá, Colombia.
- Blundo-Canto, G., Bax, V., Quintero, M., Cruz-Garcia Gisella, S., Groeneveld Rolf, A. & Perez-Marulanda, L. 2018. The Different Dimensions of Livelihood Impacts of Payments for Environmental Services (PES) Schemes: A Systematic Review. *Ecological Economics*, 149, 160-183.
- Brouwers, M. C., Kho, M. E., Browman, G. P., Burgers, J. S., Cluzeau, F., Feder, G., Fervers, B., Graham, I. D., Grimshaw, J., Hanna, S. E., Littlejohns, P., Makarski, J. & Zitzelsberger, L. 2010. AGREE II: advancing guideline development, reporting and evaluation in health care. *Cmaj*, 182, E839-42.
- Brudevold-Newman, A., Honorati, M., Jakiela, P. & Ozier, O. 2017. A Firm of One's Own: Experimental Evidence on Credit Constraints and Occupational Choice.
- Brunton, G., Oliver, S. & Thomas, J. 2020. Innovations in framework synthesis as a systematic review method. *Research Synthesis Methods*, 11, 316-330.
- Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D. & Tyrer, P. 2000. Framework for design and evaluation of complex interventions to improve health. *BMJ*, 321, 694-6.
- Carew, M., Deluca, M., Groce, N., Fwaga, S. & Kett, M. 2020. The Impact of an Inclusive Education Intervention on Learning Outcomes for Girls with Disabilities within a Resource Poor Setting, *African journal of disability*, 9, 555. <https://doi.org/10.4102/ajod.v9i0.555> .
- Catalano Richard, F., Skinner Martie, L., Alvarado, G., Kapungu, C., Reavley Nicola, J., Patton George, C., Jessee, C., Plaut, D., Moss, C., Bennett, K., Sawyer Susan, M., Sebany, M., Sexton, M., Olenik, C. & Petroni, S. 2019. Positive Youth Development Programs in Low-and Middle-Income Countries: A Conceptual Framework and Systematic Review of Efficacy, *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*, 65(1), 15–31. <https://doi.org/10.1016/j.jadohealth.2019.01.024>.
- Chirwa, E., Gelli, A., Aberman, N.-L., Margolies, A., Santacroce, M. & Baulch, B. 2017. Lean-Season Food Transfers Affect Children's Diets and Household Food

- Security: Evidence from a Quasi-Experiment in Malawi. *Journal of Nutrition*, 147, 869-878.
- Chrisp, J., Smyth, L., Stansfield, C., Pearce, N., France, R. & Taylor, C. 2022. Basic income experiments in OECD countries: A rapid evidence review. London: EPPI, Centre, Social Research Institute, UCL, UK.
- Cocciolo, S., Ghisolfi, S., Habib, A., Rashid, S. M. A. & Tompsett, A. 2020. Access to safe drinking water: experimental evidence from new water sources in Bangladesh.
- Coleman, E. A., Manyindo, J., Parker, A. R. & Schultz, B. 2019. Stakeholder engagement increases transparency, satisfaction, and civic action. *Proceedings of the National Academy of Sciences*, 116, 24486-24491.
- Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. & Petticrew, M. 2008. Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*, 337, a1655.
- Dar, M. H., Chakravorty, U. & Emerick, K. 2020. Impact of alternate wetting and drying on farm incomes and water savings in Bangladesh, h, 3ie Impact Evaluation Report 108. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: <https://doi.org/10.23846/DPW1IE108>.
- De Buck, E., Van Remoortel, H., Hannes, K., Govender, T., Naidoo, S., Avau, B., Vande Veegaete, A., Musekiwa, A., Lutje, V., Cargo, M., Mosler, H.-J., Vandekerckhove, P. & Young, T. 2017. Approaches to promote handwashing and sanitation behaviour change in low- and middle-income countries: a mixed method systematic review. *Campbell Systematic Reviews*, 13, 1-447.
- Diagne, A. & Cabral, F. J. 2017. Agricultural Transformation In Senegal: Impacts of an Integrated Program, (February 1, 2017). Partnership for Economic Policy Working Paper No. 2017-09, Available at SSRN: <https://ssrn.com/abstract=3163680> or <http://dx.doi.org/10.2139/ssrn.3163680>
- Edjekumhene, I., Voors, M., Lujala, P., Brunnschweiler, C., Owusu, C. K. & Nyamekye, A. 2019. Impacts of key provisions in Ghana's Petroleum Revenue Management Act, , 3ie Impact Evaluation Report 94. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: <https://doi.org/10.23846/TW8IE94>.
- Garn, J. V., Sclar, G. D., Freeman, M. C., Penakalapati, G., Alexander, K. T., Brooks, P., Rehfuss, E. A., Boisson, S., Medlicott, K. O. & Clasen, T. F. 2017. The Impact of Sanitation Interventions on Latrine Coverage and Latrine Use: A Systematic Review and Meta-Analysis. *International Journal of Hygiene and Environmental Health*, 220, 329-340.
- Ghose, J., Bakshi, S. K., Arora, N., Sharma, R., Deepa, N. & Govindan, M. 2017. Natural Resource Revenue Management in Low- And Middle- Income Countries Experiencing Politically Fragile Conditions: A Systematic Review, London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London. Gibbs, A., Corboz, J., Chirwa, E., Carron, M., Karim, F., Shafiq, M., Mecagni, A., Maxwell-Jones, C., Noble, E. & Jewkes, R. 2020. The impacts of combined social and economic empowerment training on intimate partner

violence, depression, gender norms and livelihoods among women: an individually randomised controlled trial and qualitative study in Afghanistan. *BMJ Global Health*, 5.

- Gough, D. 2021. Appraising Evidence Claims. *Review of Research in Education*, 45, 1-26.
- Gough, D., Kenny, C., Vigurs, C., Stansfield, C., Rosen, R. & Taylor, T. 2014. Social values related to the development of health and care guidance: literature review for NICE by its Research Support Unit. *Report of the Research Support Unit for the National Institute for Health and Care Excellence*, UK.
- Gough, D. & White, H. 2018. Evidence standards and evidence claims in web based research portals. London: Centre for Homelessness Impact, ISBN: 978-1-9995928-3-7
- Graves, J. C., Elyanu, P., Schellack, C. J., Asire, B., Prust, M. L., Prescott, M. R., Mirembe, E., Lukabwe, I., Mirembe, B., Musinguzi, J. & Moberley, S. A. 2018. Impact of a Family Clinic Day intervention on paediatric and adolescent appointment adherence and retention in antiretroviral therapy: A cluster randomized controlled trial in Uganda. *PLOS ONE*, 13, 1-18.
- Harris-Fry, H. A., Paudel, P., Harrisson, T., Shrestha, N., Jha, S., Beard, B. J., Copas, A., Shrestha, B. P., Manandhar, D. S., Costello, A. M. D. L., Cortina-Borja, M. & Saville, N. M. 2018. Participatory Women's Groups with Cash Transfers can Increase Dietary Diversity and Micronutrient Adequacy During Pregnancy, whereas Women's Groups with Food Transfers can Increase Equity in Intrahousehold Energy Allocation. *Journal of Nutrition*, 148, 1472-1483-472-1483.
- Higgins, J., Lasserson, T., Chandler, J., Tovey, D. & Churchill, R. 2018. Standards for the conduct and reporting of new Cochrane Intervention Reviews, reporting of protocols and the planning, conduct and reporting of updates. *Methodological Expectations of Cochrane Intervention Reviews (MECIR)*, Cochrane: London, February 2022.
- Hirvonen, K., Hoddinott, J., Minten, B. & Stifel, D. 2017. Children's Diets, Nutrition Knowledge, and Access to Markets. *World Development*, 95, 303-315.
- Hossain, S., Huq, N. L., Mamun, R., Nowrin, I., Ether, S., Sarker, M., Khan, A. & Adams, A. M. 2017a. What is the impact of approaches for addressing insecurity or violence arising from urbanisation? Implications of evidence for South Asia, London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- Hossain, S., Spurway, K., Zwi, A. B., Huq, N. L., Mamun, R. I. R., Nowrin, I., Ether, S., Bonnitcho, J., Dahal, N. & Adams, A. M. 2017b. What is the impact of urbanisation on risk of, and vulnerability to, natural disasters? What are the effective approaches for reducing exposure of urban population to disaster risks?, London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- Husereau, D., Drummond, M., Augustovski, F., De Bekker-Grob, E., Briggs, A. H., Carswell, C., Caulley, L.,

- Chaiyakunapruk, N., Greenberg, D., Loder, E., Mauskopf, J., Mullins, C. D., Petrou, S., Pwu, R.-F. & Staniszewska, S. 2022. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *BMJ*, 376, e067975.
- Ilavarasan, P. V., Kar, A. K. & Aswani, R. 2017a. Employment outcomes of skills training in low and middle income South Asian countries: An evidence summary, London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- Ilavarasan, P. V., Otieno, A., Huang, Y., Cabotaje, C. & Sahai, G. 2017b. Does access to business relevant information through networked devices enhance the internal efficiency and business growth of urban MSMEs in low- and middle-income countries? International Development Research Centre, Ottawa, Canada, and the Department for International Development, UK...
- Källander, K. E., Soremekun, S., Strachan, D. L. L., Hill, Z., Kasteng, F., Kertho, E., Nanyonjo, A., Ten Asbroek, G., Vassall, A., Ayebale, G., Lumumba, P., Nakirunda, M., Bagorogoza, B., Meek, S., Tibenderana, J., Lingam, R. & Kirkwood, B. 2021. Improving Community Health Worker Treatment for Malaria, Diarrhoea and Pneumonia in Uganda Through inSCALE Community and Mhealth Innovations: A Cluster Randomised Controlled Trial. *SSRN Electronic Journal*.
- Khwaja, A., Haq, O., Khan, A., Olken, B. & Shaikat, M. 2020. Rebuilding the social compact: urban service delivery and property taxes in Pakistan, n, 3ie Impact Evaluation Report 117. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: <https://doi.org/10.23846/DPW1IE117>.
- Kumar, L., Balasubramanian, V. & Raghu Raman, S. 2018. Effectiveness of Market-led Development Approaches in Low and Middle Income Countries: A Systematic Review. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London..
- Langer, L., Erasmus, Y., Tannous, N., Obuku, E., Ravat, Z., Chisoro, C., Opondo, M., Nduku, P., Tripney, J., Van Rooyen, C. & Stewart, R. 2018. Women in wage labour: A systematic review of the effectiveness and design features of interventions supporting women's participation in wage labour in higher-growth and/or male-dominated sectors in LMICs. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London ..
- Li, S. & Liu, Y. 2020. Using big data to evaluate the impacts of transportation infrastructure investment: the case of subway systems in Beijing.
- Manning, R., Goldman, I. & Hernandez Licona, G. 2020. *The impact of impact evaluation: Are impact evaluation and impact evaluation synthesis contributing to evidence generation and use in low-and middle-income countries?* WIDER Working Paper.
- Maynard, V., Parker, E. & Twigg, J. 2017. The Effectiveness and Efficiency of Interventions Supporting Shelter Self-Recovery Following Humanitarian Crises.
- McKenzie, D. 2017. Identifying and Spurring High-Growth Entrepreneurship. *American Economic Review*, 107, 2278-2307.

- Menon, K., Puthusseray, S., Ravalia, A., Panchal, P., Rana, R., Mistry, S., Tseng, P., Bhandol, J. & Mavalankar, D. 2018. Effectiveness of nutrition interventions in low and middle income countries: An evidence summary. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M. & Altman, D. G. 2012. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *International Journal of Surgery*, 10, 28-55.
- Morrow, S.L. 2005. Quality and trustworthiness in qualitative research in counseling psychology. *Journal of Counseling Psychology*, 52, 250-260.
- Morten, M., Bryan, G., Siddiqi, B. & Balboni, C. 2020. Evaluating the impacts of the Dar es Salaam Bus Rapid Transit System, 3ie Impact Evaluation Report 110. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: <https://doi.org/10.23846/DPW11E110>.
- Mvukiyehe, E. & van der Windt, P. 2020. Assessing the Longer Term Impact of Community-driven Development Programs: Evidence from a Field Experiment in the Democratic Republic of Congo, Policy Research Working Papers No. 9140. World Bank, Washington, DC. <https://openknowledge.worldbank.org/entities/publication/d6f5a768-a89b-5a1c-9ff6-80f95263156d> License: CC BY 3.0 IGO
- Nair, N. S., Darak, S., Parsekar, S. S., Parchure, R., Darak, T., Nelson, H., Vijayamma, R. & Menon, S. 2017a. Effectiveness of Behaviour Change Communication interventions in improving the delivery of health messages for ante-natal care in limited literacy settings: An evidence summary. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- Nair, N. S., Shrinivas, D., Bhumika, T. V., Trupti, D., Mathews, M., Dayashwori, D., Ratheebhai, V. & Anjali, D. 2017b. Gender-responsive policing initiatives designed to enhance confidence, satisfaction in policing services and reduce risk of violence against women in low and middle income countries – A systematic review. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London,.
- Nair, N. S., Venkatesh, B. T., Bhageerathy, R. K. A., Unnikrishnan, B., Sindhu, S. U., Pundir, P., Sebastian, A. M., Krishnan, J. B., Guddattu, V., Vijayamma, R., Das, U. & Sequeria, S. M. 2017c. Public works programmes: How effective are public works programmes in stimulating local economic transformation in low and middle income countries? A systematic review. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- Najy, B., David, M., Victor, P. & Massimiliano, S. 2018. Does Inducing Informal Firms to Formalize Make Sense? Experimental Evidence from Benin. *Journal of Public Economics*, 157, 1-14.

- Nidhi, S., Neha, B., Bhawna, T., Sonakshi, S., Bhattacharya, P. K., Shazia, A. & Veena, A. 2017. Effects of Various Disaster Management Approaches: An Evidence Summary. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London K.
- O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A. & Cook, D. A. 2014. Standards for Reporting Qualitative Research: A Synthesis of Recommendations. *Academic Medicine*, 89.
- Obuku, E. A., Stewart, R., Mijumbi, R., Ochan, M., Achana, F., Akena, D., Nakitende, A., Ssemata, A., Kinengyere, A., Semakula, D., Ssenono, R., Nsangi, A., Lalitha, R., Mwesiga, E., Akite, J., Basaza, R., Newbrander, W., Okello, J. & Sewankambo, N. 2017. Working with non-state providers in post-conflict and fragile states in primary healthcare service delivery: a systematic review, London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London..
- Oliver, S., Mincheko, D., Bangpan M., Dickson K., Stansfield, C., Tripney J. (2023) 'Evidence claims for informing decisions relating to socio-economic development'. CEDIL Lessons Learned Paper 2, CEDIL, Oxford. Available at <https://doi.org/10.51744/LLP2>
- Oliver, S., Harden, A., Rees, R., Shepherd, J., Brunton, G., Garcia, J. & Oakley, A. 2005. An emerging framework for including different types of evidence in systematic reviews for public policy. *Evaluation*, 11, 428-446.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P. & Moher, D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71.
- Parker, R., Coleman, E., Manyindo, J., Mukuru, E. & Schultz, B. 2019. Impacts of Community Stakeholder Engagement Interventions in Ugandan Oil Extractives.
- Patel, R., King, J., Phelps, L. & Sanderson, D. 2017. What Practices are Used to Identify and Prioritize Vulnerable Populations Affected by Urban Humanitarian Emergencies? *Oxfam GB*, 1-50.
- Pellegrini, L. 2018. Behavioural responses to information on contaminated drinking water: randomized evidence from the Ecuadorian Amazon, , 3ie Grantee Final Report. New Delhi: International Initiative for Impact Evaluation (3ie).
- Pilkington, G., Panday, S., Khatib, M. N., Kotas, E., Hill, R. A., Simkhada & Jones, L. 2017. The effectiveness of community engagement and participation approaches in low and middle income countries: a review of systematic reviews with particular reference to the countries of South Asia, London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London..
- Pinnock, H., Barwick, M., Carpenter, C. R., Eldridge, S., Grandes, G., Griffiths, C. J., Rycroft-Malone, J., Meissner, P., Murray, E., Patel, A., Sheikh, A. & Taylor, S. J. C.

2017. Standards for Reporting Implementation Studies (StaRI) Statement. *BMJ*, 356, i6795.
- Piper, B., Simmons Zuilkowski, S., Dubeck, M., Jepkemei, E. & King, S. J. 2018. Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers' guides. *World Development*, 106, 324-336.
- Rakotonarivo, O. S., Jacobsen, J. B., Larsen, H. O., Jones, J. P., Nielsen, M. R., Ramamonjisoa, B. S., Mandimbinaina, R. H. & Hockley, N. 2017. Qualitative and quantitative evidence on the true local welfare costs of forest conservation in Madagascar: Are discrete choice experiments a valid ex ante tool? *World Development*, 94, 478-491.
- Robertson, R. 2019. Lights On: How Transparency Increases Compliance in Cambodian Global Value Chains. *ILR Review*, 73, 1-30.
- Roth, V., Araar, A., Sry, B. & Phann, D. 2017. The Dynamics of Microcredit Borrowings in Cambodia, (March 1, 2017). Partnership for Economic Policy Working Paper No. 2017-17, Available at SSRN: <https://ssrn.com/abstract=3163570> or <http://dx.doi.org/10.2139/ssrn.3163570>
- Saboya, M., Rudiger, J., Frize, J., Ruegenberg, D., Seco, A. R. & McMillon, C. 2018. Impact Evaluation of WFP's Programs Targeting Moderate Acute Malnutrition in Humanitarian Situations in Chad, . 3ie Impact Evaluation Report 78. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: doi: <https://doi.org/10.23846/tw6IE78>.
- Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E. & Henry, D. A. 2017. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358, j4008.
- Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M. & Moore, L. 2021. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ*, 374, n2061.
- Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T. & Boutron, I. 2016. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355.
- Sterne, J. A., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S. & Eldridge, S. M. 2019. RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, 366.
- Stone, R., De Hoop, T., Coombes, A. & Nakamura, P. 2020. What works to improve early grade literacy in Latin America and the Caribbean? A systematic review and meta-analysis. *Campbell Systematic Reviews*, 16, e1067.
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., Moher, D., Becker, B. J., Sipe, T. A., Thacker, S. B. 2000. Meta-analysis of Observational Studies in Epidemiology A Proposal for Reporting. *JAMA*, 283, 2008-2012.

- Tong, A., Sainsbury, P. & Craig, J. 2007. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 19, 349-357.
- Tsai, L. L., McMurry, N. & Rajeswaran, S. 2018. The Effect of Civic Leadership Training on Citizen Engagement and Government Responsiveness: Experimental Evidence from the Philippines, Making All Voices Count Research Report, Brighton: IDS.
- Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C. & Vandenbroucke, J. P. 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *Annals of Internal Medicine*, 147, 573-577.
- Waddington, H., Sonnenfeld, A., Finetti, J., Gaarder, M., John, D. & Stevenson, J. 2019. Citizen engagement in public services in low- and middle-income countries: A mixed-methods systematic review of participation, inclusion, transparency and accountability (PITA) initiatives. *Campbell Systematic Reviews*, 15, e1025.
- Whiting, P., Savović, J., Higgins, J. P., Caldwell, D. M., Reeves, B. C., Shea, B., Davies, P., Kleijnen, J. & Churchill, R. 2016. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, 69, 225-234.
- Williamson, K., Landis, D., Shannon, H., Gupta, P. & Gillespie, L.-A. 2017. The Impact of Protection Interventions on Unaccompanied and Separated Children in Humanitarian Crises, Humanitarian Evidence Programme. Oxford: Oxfam GB.
- Wolf, J., Hunter, P. R., Freeman, M. C., Cumming, O., Clasen, T., Bartram, J., Higgins, J. P. T., Johnston, R., Medlicott, K., Boisson, S. & Prüss-Ustün, A. 2018. Impact of Drinking Water, Sanitation and Hand Washing with Soap on Childhood Diarrhoeal Disease: Updated Meta-Analysis and Regression. *Tropical Medicine and International Health*, 23, 508-525.
- Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J. & Pawson, R. 2013. RAMESES publication standards: realist syntheses. *BMC Medicine*, 11, 21.
- Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J. & Greenhalgh, T. 2016. RAMESES II reporting standards for realist evaluations. *BMC Medicine*, 14, 96.
- Yount, K. M., Krause, K. H. & Miedema, S. S. 2017. Preventing Gender-Based Violence Victimization in Adolescent Girls in Lower-Income Countries: Systematic Review of Reviews. *Social Science & Medicine*, 192, 1-13.
- Zegarra, E., Higuchi, A. & Vargas, R. 2017. Assessing the impacts of a training program for women in Peru: Are there social networking effects? Partnership for Economic Policy Working Paper No. 2017-02, Available at SSRN: <https://ssrn.com/abstract=3164295> or <http://dx.doi.org/10.2139/ssrn.3164295>
- Zwi, A. B., Spurway, K., Marincowitz, R., Ranmuthugala, G., Hobday, K. & Thompson, L. 2018. Do CBDRM initiatives impact on the social and economic costs of disasters? If so, how, why, when and in what way(s)? London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

References from Part B

- ALSHARARI NIZAR, M. & AL-SHBOUL, M. 2019. Evaluating qualitative research in management accounting using the criteria of "convincingness". *Pacific Accounting Review*, 31, 43-62.
- ANTONAKIS, J., BENDAHAN, S., JACQUART, P. & LALIVE, R. 2010. On making causal claims: A review and recommendations. *Leadership Quarterly*, 21, 1086-1120.
- ATTANASIO, O. & CAVATORTA, E. 2017. Complements and Alternatives to RCTs in Evaluation in Development. *CEDIL Pre-Inception Paper*. London: Centre of Excellence for Development Impact and Learning (CEDIL).
- AVIS, M. 1995. Valid arguments? A consideration of the concept of validity in establishing the credibility of research findings. *Journal of Advanced Nursing*, 22, 1203-1209.
- BERO, L., CHIU, K. & GRUNDY, Q. 2019. The SSSPIN study-spin in studies of spin: meta-research analysis. *BMJ (Clinical research ed.)*, 367, l6202.
- BLEAKLEY, C., REIJGERS, J. & SMOLIGA, J. M. 2020. Many High-Quality Randomized Controlled Trials in Sports Physical Therapy Are Making False-Positive Claims of Treatment Effect: A Systematic Survey. *Journal of Orthopaedic & Sports Physical Therapy*, 50, 104-109.
- BOUTRON, I., DUTTON, S., RAVAUD, P. & ALTMAN, D. G. 2010. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*, 303, 2058-64.
- BOUTRON, I. & RAVAUD, P. 2018. Misrepresentation and distortion of research in biomedical literature. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 2613-2619.
- CARTWRIGHT, N. 2021. Rigour versus the need for evidential diversity. *Synthese*, 199, 13095-13119.
- CARTWRIGHT, N., CHARLTON, L., JUDEN, M., MUNSLOW, T. & WILLIAMS, R. B. 2020. Making predictions of programme success more reliable. *CEDIL Methods Working Paper*. London/Oxford: Centre of Excellence for Development Impact and Learning (CEDIL).
- CARTWRIGHT, N., GOLDFINCH, A. & HOWICK, J. 2010. Evidence-based policy: where is our theory of evidence? *Journal of Children's Services*, 4, 6-14.
- CHIU, K., GRUNDY, Q. & BERO, L. 2017. 'Spin' in published biomedical literature: A methodological systematic review. *PLoS biology*, 15, e2002173.
- CHRISP, J., SMYTH, L., STANSFIELD, C., PEARCE, N., FRANCE, R. & TAYLOR, C. 2022. Basic income experiments in OECD countries: A rapid evidence review. London: EPPI, Centre, Social Research Institute, UCL.
- CLAVEAU, F. 2011. Evidential variety as a source of credibility for causal inference: beyond sharp designs and structural models. *Journal of Economic Methodology*, 18, 233-253.
- DAVIDSON, E. J. 2014. *Evaluative Reasoning*. Florence Italy: UNICEF Office of Research.

- DELAHAIS, T. & TOULEMONDE, J. 2017. Making rigorous causal claims in a real-life context: Has research contributed to sustainable forest management? *Evaluation*, 23, 370-388.
- DENNING PETER, J. 2011. The grounding practice. *Communications of the ACM*, 54, 38-40.
- DIONG, J., BUTLER, A. A., GANDEVIA, S. C. & HÉROUX, M. E. 2018. Poor statistical reporting, inadequate data presentation and spin persist despite editorial advice. *PloS one*, 13, e0202121.
- EVERS, C. W. & MASON, M. 2011. Context based inferences in research methodology: the role of culture in justifying knowledge claims. *Comparative Education*, 47, 301-314.
- FARLEY, J. 1987. Justifying conclusions in naturalistic evaluations: A critical perspective. *Evaluation and Program Planning*, 10, 343-350.
- GHANNAD, M., OLSEN, M., BOUTRON, I. & BOSSUYT, P. M. 2019. A systematic review finds that spin or interpretation bias is abundant in evaluations of ovarian cancer biomarkers. *Journal of clinical epidemiology*, 116, 9-17.
- GILLESPIE, D. L. & LEFFLER, A. 1987. The Politics of Research Methodology in Claims-Making Activities - Social-Science and Sexual Harassment. *Social Problems*, 34, 490-501.
- GLINER JA (1994) Reviewing qualitative research: proposed criteria for fairness and rigour, *Occupational Therapy Journal of Research*, 14(2), 78-92.
- GORARD, S. & TAN, Y. 2021. Reviewing the logic of social scientific claims. *Technium Social Sciences Journal*, 24, 113-124.
- GORARD, S. & TAN, Y. 2022. The difficulty of making claims to knowledge in social science. *Technium Social Sciences Journal*, 170-202.
- GOUGH, D. 2021. Appraising Evidence Claims. *Review of Research in Education*, 45, 1-26.
- GOUGH, D. & WHITE, H. 2018. Evidence standards and evidence claims in web based research portals. *London: Centre for Homelessness Impact*.
- GRAESSER ARTHUR, C. & XIANGEN, H. 2011. Commentary on Causal Prescriptive Statements. *Educational Psychology Review*, 23, 279-285.
- GREENE JENNIFER, C. 1987. Justifying conclusions in naturalistic evaluations: A practical perspective. *Evaluation and Program Planning*, 10, 325-333.
- HAIYAN, B. 2011. Using Propensity Score Analysis for Making Causal Claims in Research Articles. *Educational Psychology Review*, 23, 273-278.
- HILTON, L. & JONAS, W. B. 2017. Claim Assessment Profile: A Method for Capturing Healthcare Evidence in the Scientific Evaluation and Review of Claims in Health Care (SEaRCH). *Journal of Alternative and Complementary Medicine*, 23, 96-+.
- IOANNIDIS, J. P. 2005. Why most published research findings are false. *PLoS medicine*, 2, e124.
- ITO, C., HASHIMOTO, A., UEMURA, K. & OBA, K. 2021. Misleading Reporting (Spin) in Noninferiority Randomized Clinical Trials in Oncology With Statistically Not Significant Results: A Systematic Review. *JAMA network open*, 4, e2135765.

- JONAS WAYNE, B., CRAWFORD, C., HILTON, L. & ELFENBAUM, P. 2017. Scientific evaluation and review of claims in health care (SEaRCH): A streamlined, systematic, phased approach for determining "what works" in healthcare. *The Journal of Alternative and Complementary Medicine*, 23, 18-25.
- KEMPF, E., DE BEYER, J. A., COOK, J., HOLMES, J., MOHAMMED, S., NGUYÊN, T. L., SIMERA, I., TRIVELLA, M., ALTMAN, D. G., HOPEWELL, S., MOONS, K. G. M., PORCHER, R., REITSMA, J. B., SAUERBREI, W. & COLLINS, G. S. 2018. Overinterpretation and misreporting of prognostic factor studies in oncology: a systematic review. *British journal of cancer*, 119, 1288-1296.
- KULIKOWICH JONNA, M. & SPERLING RAYNE, A. 2011. Prescriptive Statements: Philosophical, Theoretical, and Methodological Considerations. *Educational Psychology Review*, 23, 189-195.
- KUORIKOSKI, J. & MARCHIONNI, C. 2016. Evidential diversity and the triangulation of phenomena. *Philosophy of Science*, 83, 227-247.
- LAZARUS, C., HANEEF, R., RAVAUD, P. & BOUTRON, I. 2015. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. *BMC Medical Research Methodology*, 15, 85.
- LEVIN JR (2004) Random thoughts on the (in)credibility of educational psychological intervention research, *Educational Psychologist*, 39(3). 173-184.
- MARLEY SCOTT, C. & LEVIN JOEL, R. 2011. When Are Prescriptive Statements in Educational Research Justified? *Educational Psychology Review*, 23, 197-206.
- MARROW SL (2005) Quality and trustworthiness in qualitative research in counseling psychology. *Journa of Counseling Psychology*, 52(2), 250-60.
- MCCLINTOCK, C. 1987. Administrators as information brokers: A managerial perspective on naturalistic evaluation. *Evaluation and Program Planning*, 10, 315-323.
- MCGRATH, T. A., MCINNES, M. D. F., VAN ES, N., LEEFLANG, M. M. G., KOREVAAR, D. A. & BOSSUYT, P. M. M. 2017. Overinterpretation of Research Findings: Evidence of "Spin" in Systematic Reviews of Diagnostic Accuracy Studies. *Clinical chemistry*, 63, 1353-1362.
- MINGERS, J. & STANDING, C. 2020. A Framework for Validating Information Systems Research Based on a Pluralist Account of Truth and Correctness. *Journal Of The Association For Information Systems*, 117-151.
- MORAN-ELLIS, J., ALEXANDER, V. D., CRONIN, A., DICKINSON, M., FIELDING, J., SLENEY, J. & THOMAS, H. 2016. Triangulation and integration: processes, claims and implications. *Qualitative Research*, 6, 45-59.
- MORTON, A. 2011. Conventional Norms of Reasoning. *Dialogue*, 50, 247-260.
- NOLEN, A. & TALBERT, T. 2011. Qualitative Assertions as Prescriptive Statements. *Educational Psychology Review*, 23, 263-271.
- OAKSFORD, M., CHATER N. (2020) New paradigms in the Psychology of reasoning, *Annual Review of Psychology*, 71(1): 305-30.

- O'CONNELL ANN, A. & GRAY DELEON, L. 2011. Cause and Event: Supporting Causal Claims through Logistic Models. *Educational Psychology Review*, 23, 245-261.
- OCHODO, E. A., DE HAAN, M. C., REITSMA, J. B., HOOFT, L., BOSSUYT, P. M. & LEEFLANG, M. M. 2013. Overinterpretation and misreporting of diagnostic accuracy studies: evidence of "spin". *Radiology*, 267, 581-8.
- OLARTE PARRA, C., BERTIZZOLO, L., SCHROTER, S., DECHARTRES, A. & GOETGHEBEUR, E. 2021. Consistency of causal claims in observational studies: a review of papers published in a general medical journal. *BMJ open*, 11, e043339.
- ORMEROD RJ (2010) Rational inference: Deductive. Inductive and Probabilistic Thinking, *Journal of the Operational Research Society*, 61: 1207-1223.
- OSIMANI B., LANDES, J. (2021) Varieties of Error and Varieties of Evidence in Scientific Inference, *The British Journal for the Philosophy of Science*
- PEARL J. (2010) Causality in the social and behavioural sciences, UCLA Cognitive Systems Laboratory, Technical Report, UCLA, USA.
- PEARSOL JAMES, A. 1987. Justifying conclusions in naturalistic evaluations: An interpretive perspective. *Evaluation and Program Planning*, 10, 335-341.
- PETERS, U. 2020. Values in science: assessing the case for mixed claims. *Inquiry: An Interdisciplinary Journal Of Philosophy*.
- REHG, W. & STALEY, K. 2008. The CDF collaboration and argumentation theory: The role of process in objective knowledge. *Perspectives on science*, 16, 1-25.
- SHIRVANI, S., RIVES-LANGE, C., RASSY, N., BERGER, A., CARETTE, C., POGHOSYAN, T. & CZERNICHOW, S. 2022. Spin in the Scientific Literature on Bariatric Endoscopy: a Systematic Review of Randomized Controlled Trials. *Obesity surgery*, 32, 503-511.
- SHUYAN, S. & WEI, P. 2011. The Philosophical Foundations of Prescriptive Statements and Statistical Inference. *Educational Psychology Review*, 23, 207-220.
- SHYAGALI TARULATHA, R., RATHORE, A., KAPOOR, S., GUPTA, A., TIWARI, A. & PATIDAR, R. 2022. Evaluation of exaggerated claims in the abstracts of systematic reviews reporting accelerated orthodontic tooth movement: a meta research analysis. *Australasian Orthodontic Journal*, 38, 268-277.
- SMITH JOHN, K. 1987. Commentary: Relativism and justifying conclusions in naturalistic evaluations. *Evaluation and Program Planning*, 10, 351-358.
- SMITH, N. L. 1987. Toward the Justification of Claims in Evaluation Research. *Evaluation and Program Planning*, 10, 309-314.
- SNODGRASS MELINDA, R., CHUNG MOON, Y., KRETZER JAMES, M. & BIGGS ELIZABETH, E. 2022. Rigorous Assessment of Social Validity: A Scoping Review of a 40-Year Conversation. *Remedial And Special Education*, 43, 114-130.
- STALEY KENT, W. 2004. Robust Evidence and Secure Evidence Claims. *Philosophy of Science*, 71, 467-488.
- STEPHEN, J. 2021. Scientific deceit. *Synthese*, 373-394.
- SUN W, PAN S (2011) The Philosophical Foundations of Prescriptive Statements and

- Statistical Inference. *Educational Psychology Review* 23: 207-220.
- THAPA, D. K., VISENTIN, D. C., HUNT, G. E., WATSON, R. & CLEARLY, M. 2020. Being honest with causal language in writing for publication. *Journal of advanced nursing*, 76, 1285-1288.
- VELDE, H. M., VAN HETEREN, J. A. A., SMIT, A. L. & STEGEMAN, I. 2021. Spin in Published Reports of Tinnitus Randomized Controlled Trials: Evidence of Overinterpretation of Results. *Frontiers in neurology*, 12, 693937.
- VERMAAS, P. E. 2016. Towards Precedence that Justifies the Knowledge Claims of Design Methods. *Design Journal*, 19, 195-204.
- WATKINS TIMOTHY, R. 2018. *Understanding uncertainty and bias to improve causal inference in health intervention research*, PhD Dissertation, University of Sydney, Australia.
- WECKER, C. 2013. How to Support Prescriptive Statements by Empirical Research: Some Missing Parts. *Educational Psychology Review*, 25, 1-18.
- WELLMER, A. 1993. Truth, Contingency and Modernity + Justifying Truth Claims Using Standards of Argumentation or Evidential Support. *Modern Philology*, 90, S109-S124.
- WILLIAMS DAVID, D. 1986. Naturalistic Evaluation: Potential Conflicts Between Evaluation Standards and Criteria for Conducting Naturalistic Inquiry. *Educational Evaluation And Policy Analysis*, 8, 87-99.
- XIONG, C., ZHAN, Y. & CHEN, S. 2015. An argumentation model based on evidence theory. *International Conference On Computer Science And Education*.
- YAVCHITZ, A., RAVAUD, P., ALTMAN, D. G., MOHER, D., HROBJARTSSON, A., LASSERSON, T. & BOUTRON, I. 2016. A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity. *Journal of clinical epidemiology*, 75, 56-65.
- YI, S. & DUVAL-COUEIL, N. 2022. Standards for Evaluating Impact in Entrepreneurship Education Research: Using a Descriptive Validity Framework to Enhance Methodological Rigor and Transparency. *Entrepreneurship Theory And Practice*, 46, 1685-1716.
- YUSSEN STEVEN, R. 2011. When Is It Acceptable to Make Prescriptive Statements in Educational Research Articles? Commentary on the Special Issue Papers. *Educational Psychology Review*, 23, 287-292.
- ZALANGA, S. 2011. The Use of Empiricist and Narrative Methods in Comparative Social Science Research: Lessons and Insights for Third World Studies and Research. *The Journal Of Third World Studies*, 28, 107.



 @CEDILProgramme

 Centre of Excellence for Development Impact

<https://cedilprogramme.org>