



CEDIL LESSONS LEARNED PAPER 2

# Evidence claims for informing decisions relating to socio-economic development

Sandy Oliver, Dayana Minchenko, Mukdarut Banpgan, Kelly Dickson, Claire Stansfield, and Janice Tripney



## 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 Lessons Learned Paper

The CEDIL Lessons Learned series shares learning from across the CEDIL programme.

## About this working paper

Please direct any comments or queries to the corresponding author, Prof Sandy Oliver, at [sandy.oliver@ucl.ac.uk](mailto:sandy.oliver@ucl.ac.uk)

**Suggested citation:** 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>

Cover design: PhilDoesDesign

Cover images: (clockwise from top left) UN Women Asia and the Pacific; Kyaw Win Khaing, WorldFish; Sergio Hola; UN Women/Ryan Brown, all shared on flickr under Creative Commons licences

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

# Evidence claims for informing decisions relating to socio-economic development

## Authors

Sandy Oliver

Dayana Minchenko

Mukdarut Bangpan

Kelly Dickson

Claire Stansfield

Janice Tripney

## Institute

EPPI-Centre, UCL Social Research Institute





---

## Contents

Evidence claims for informing decisions relating to socio-economic development .....	3
Glossary of terms and abbreviations used in the working paper .....	6
Abstract.....	9
Background .....	9
Aim of the work.....	9
Section 1 .....	12
Introduction.....	12
Section 2 .....	13
Research problem and questions .....	13
Section 3 .....	14
Study design and methods .....	14
Data we analysed .....	16
Data analysis .....	18
Section 4 .....	19
Phase I: Developing an evidence claims framework.....	19
Phase I: Current debates highlighted by the workshop .....	23
Phase I: Making and justifying claims in CEDIL design papers .....	27
Phase I: Communicating claims for other audiences.....	31
Phase I: Discussion .....	33
Section 5 .....	38
Phase II: Applying an evidence claims framework.....	38
Phase II: Focus of claims about developing effective complex interventions	38
Phase II: Scope of claims, in terms of certainty and reach.....	40
Phase II: Backdrop for claims, in terms of their framing and fitness for purpose..	41
Phase II: Discussion.....	42
Section 6 .....	44
Summary, conclusions and recommendations .....	44
References.....	45
CEDIL Programme references .....	45
Other references.....	48

---

Acknowledgements .....	51
Appendix 1 .....	52
Invitation to workshop.....	52
Appendix 2 .....	56
Outputs from CEDIL’s three programmes of work.....	56
Empirical claims made by CEDIL studies.....	72
Methodological claims made by CEDIL studies.....	87
Communicating claims made by CEDIL studies.....	99
Disability Evidence Portal – Rating System Appraisal Checklist.....	105

---

## List of tables, figures, and boxes

Table 1:	Empirical and methodological aims of Design Papers	19
Box 1:	Discussants at workshop	25
Figure 1:	Framework for developing and evaluating complex interventions	29
Figure 2:	Framework for developing and evaluating research methods	31
Figure 3:	Framework for developing, justifying and communicating claims	37

---

## Glossary of terms and abbreviations used in the working paper

**CEDIL:** Centre of Excellence for Development Impact and Learning – an initiative supported by aid from the UK Government that is designed to develop and promote new impact evaluation methods for use in international development.

**Consultation:** A method of engaging citizens that has historically been viewed as offering limited power to citizens to influence research and that is often considered to be tokenistic (Tritter and McCallum, 2006).

**Context:** Refers to the circumstances or setting in which a piece of research or a phenomenon takes place. Depending on the phenomenon being studied, and the way in which it is being studied, the context of interest can vary. For example, if we were conducting a single case study of maternal feeding practices, the context that would be of most interest might be the immediate household environment, with other elements of context (such as the village or wider family) also potentially of interest. In contrast, if we were conducting a survey of infant feeding practices across settings, several different levels of context (including the country level) may be of interest.

**Decision maker:** A term used in this research to denote a particular type of stakeholder. A decision maker sets the policy within the context of the research, including the policy around the design, funding and implementation of services and/or interventions. For example, in a piece of research about adolescent contraception in a given area, decision makers could include those who fund sexual health clinics, those who develop policies around where sexual health clinics should be located, and those who develop policies around which services should be offered at sexual health clinics, and how those services should be delivered.

**Design papers:** CEDIL design papers present the protocols of studies supported by the CEDIL programme. They describe the objectives, the methodology and the implementation plan of impact evaluations, evidence syntheses, and methodological papers supported by CEDIL.

**Equity:** Refers to the absence of differences between individuals, entities and populations that are known to be unnecessary and avoidable, and which emerge through processes and practices that are unfair and unjust (Welch *et al.*, 2012).

**Evidence:** In a broad sense, evidence refers to facts or testimony in support of (or in opposition to) a conclusion, view, statement or belief (Rychetnik *et al.*, 2004). What counts as 'evidence' varies substantially, depending on the context and the question being asked. Knowledge, on the other hand, is a justified belief that can be derived from interpretations of evidence, or from practice or experience.

**Evidence claim:** A statement that is said to be, or that it is implied is, supported by research evidence, although other people may question whether the evidence supporting it has been compiled or scrutinised appropriately.

**FCDO:** The Foreign, Commonwealth and Development Office – a UK government department that now includes all international development functions of the UK Government.

**Generalisable knowledge:** Knowledge claims that rest on explicit, codified knowledge that can be widely transferred through statistical or theoretical inference, and that are developed through primary studies and research syntheses to offer new conceptualisations, theoretical understanding or empirical evidence. The kinds of generalisable knowledge that are useful for development policy include knowledge about the nature and scale of social concerns, and knowledge that recognises and explains causal relationships, including mediators and moderators. Knowledge generated in one setting is generalisable to other settings if the influence of contextual factors is either insignificant or well-understood.

**GESI:** Gender equity and social inclusion.

**ICA:** Intervention component analysis.

**Local knowledge or context-specific knowledge:** Knowledge claims that rest on explicit, codified knowledge developed through local primary studies. Also, knowledge claims that rest on familiarity with local settings, cultures and politics to offer the following: tacit understanding of the nature and scale of local issues (where ‘scale’ may rest on impressions indicating matters of priority, rather than accurate measures); recognition of trends and forecasting (essential for planning); practice and organisational ‘know-how’; and sensitivity to context that is essential for considering the appropriateness of interventions, and for gaining insights into the transferability of evidence of effects. On the one hand, evidence synthesis offers a method for providing knowledge claims that are expected to hold widely, thereby creating knowledge from and for widespread groups; and, on the other hand, it can inform deliberation by specific groups to integrate generalisable knowledge with knowledge of their local context.

**Logic model:** A graphical representation of intervention processes, and outcomes, linked by arrows indicating the direction of effect, which are developed into chains of cause-and-effect relationships.

**LMIC:** Low- or middle-income country, as defined by the World Bank.

**Positive deviance inquiry:** An approach for exploring the factors that can explain why some individuals or other social entities unexpectedly achieve desired outcomes. For example, positive deviance inquiry might explore why some children grow and develop well in otherwise harsh environments (Lapping *et al.*, 2002).

**PoW:** Programme of work

**QCA:** Qualitative comparative analysis

**RCT:** Randomised controlled trial

**Stakeholder:** A seemingly innocuous term which is often contested. In research, stakeholders are those individuals with a contract, claim, obligation, duty or responsibility relating to, or a stake in, any part of the research process. They can be

defined as 'organizations and individuals that are involved in a specific activity because they participate in producing, consuming, managing, regulating, or evaluating the activity' (Hyder *et al.*, 2010). They may take different roles, such as funders, influencers, collaborators, recipients or beneficiaries of research. The term stakeholder is often used to denote those who indirectly or directly impact the research, or who are impacted by the findings of the research or the challenge which the research seeks to explore or address.

**Stakeholder engagement** (or stakeholder involvement): Used to describe a number of different activities (many of which are described in this working paper) that seek to ensure that the critical insights of stakeholders inform all aspects of the research process, from design to data collection, to analysis and interpretation, to dissemination. Stakeholder engagement activities differ according to several different axes. One way of understanding different forms of stakeholder engagement activities is to identify the extent to which stakeholder insights inform the conduct of research (i.e. where the balance of power lies between researcher and stakeholder). The weakest levels of influence are where stakeholders are informed about research and are a passive audience to whom the findings are disseminated; in contrast, a co-production model rests on the principle of equal partnership for equal benefit.

**System:** Social systems are a set of interrelationships between individuals, groups and institutions that form a coherent whole that is complex and adaptive. 'Systems are dynamic and constantly changing; systems themselves exist within other, interdependent systems (e.g. individual, organisation, community); changes in one part of the system can have unexpected changes in other parts of the system' (Best and Holmes, 2010, p. 148).

**Systematic review:** Systematic reviews aim to identify as much as possible of the research that is relevant to the particular research questions, and use explicit methods to identify what can reliably be said on the basis of these studies (see: <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=67>).

**Theory of change:** Like logic models, theories of change are used to graphically represent complex interventions. Unlike logic models, theories of change are more explanatory, as they require all of the underlying assumptions of how and why different components, activities and outputs lead to a change in outcomes to be hypothesised.

**Uncertainty:** In terms of generalisable knowledge, uncertainty reflects the accuracy, precision or meaning of the research findings or the underlying key concepts. In terms of local knowledge, uncertainty reflects the context in which decisions are to be applied, where there may be varying levels of consensus, familiarity and predictability.

**WASH:** Water, sanitation and hygiene.



## Abstract

### Background

The Centre of Excellence for Development Impact and Learning (CEDIL) develops and tests innovative methods for evaluation and evidence synthesis. Claims made in CEDIL studies are intended to inform socio-economic development in low- and middle-income countries (LMICs), or research about LMICs. This paper provides an overview of how, in CEDIL-funded studies, claims arising from research (termed 'evidence claims' for brevity) have been justified and communicated in order to inform policy decisions relating to socio-economic development.

### Aim of the work

This study addresses two important questions about research that is designed to produce findings for decision makers. First, how are the claims that arise from the findings justified? In other words, what are the criteria (evidence standards) that studies use, explicitly or implicitly, in order to justify their evidence claims? Second, how are those claims and justifications communicated in original research reports and other outputs that share the research more widely?

### Methods

This study had two phases. In Phase I, we compiled and analysed debates about, and research practices relating to, evidence claims. Data came from the following sources: CEDIL guidance; discussions at a workshop convened by CEDIL to discuss evidence to enable policy, practice and decision-making; and evidence claims in early outputs from empirical studies conducted as part of CEDIL's research programme focusing on evaluating complex interventions, enhancing evidence transferability, and increasing evidence use. These early outputs were study design papers and communications aimed at wider audiences. Phase II analysed the final reports of CEDIL studies to understand how evidence claims were justified in practice.

### Findings

**In Phase I** of the study, the discussions during the CEDIL-convened workshop encouraged the research community : to focus on the demand for evidence to design research that strengthens confidence in the findings that matter most for decision-making; to promote the uptake of methodological innovations through professional evaluation associations; and to invite wider peer review to produce greater clarity of synthesis and interpretation.

The document review identified various types of empirical and methodological claims, in the CEDIL studies that were analysed (including systematic reviews and single studies), about developing and implementing interventions (not only about their effects), and about the influence of contextual factors. In these reviews and studies, empirical claims resting on well-established methodologies tended to be implicit. The breadth of claims is reflected in an established conceptual framework for developing and evaluating complex interventions for health and social care services, public health practice, and

other areas of social and economic policy. In contrast, the authors of the CEDIL studies we analysed supported their claims of methodological innovation more explicitly through providing detailed arguments, by referring to 'proof of concept' studies, or by citing other methodological studies. The well-established framework for developing and evaluating complex interventions can be re-cast in an analogous conceptual framework for developing and evaluating research and evaluation methods.

In the reviewed CEDIL studies and reviews, explicit descriptions of innovative methods underpinning empirical claims were essentially claims regarding methodological innovation, and an initial step for their further development and wider use within the research community. Some of the CEDIL study authors communicated evidence claims to wider audiences, visually and through blogs, which is a practice that may motivate and facilitate the use of evidence for decision-making.

Phase I ended with merging the two conceptual frameworks to create a framework for developing, justifying and communicating claims, including both empirical claims and methodological claims.

In Phase II, applying this new framework to CEDIL study reports confirmed that opaque, changing and unpredictable contexts present major challenges to longitudinal prospective studies (typically randomised controlled trials (RCT)) as regards making claims that attribute causality. In the CEDIL study reports we analysed, such challenges were overcome in part by drawing on the knowledge of diverse stakeholders to understand the contexts where they live and work, and to develop theories about how interventions may play out in particular contexts. Other partial solutions included collecting data remotely (for instance, via satellite or mobile phones), analysing vast amounts of data using machine learning or artificial intelligence, and conducting multidimensional analysis of population diversity and equity. Although methods are advancing rapidly in these two areas, many of the CEDIL studies we analysed paid scant attention to social norms or the complexities of diversity.

## **Conclusions**

Implicit evidence claims that rest on well-established methods may well cluster around fields that have attracted academic attention for some time. Explicit claims that are justified by arguments supporting the novel methods employed may well cluster around the more challenging policy fields of under-resourced or fragile settings that have a shorter history of receiving academic attention.

Widely accepted research standards encourage evaluations which demonstrate internal validity. Such standards are not necessarily apparent in the CEDIL study reports that we analysed. Moreover, the standards themselves pay little attention to the reporting of study contexts, particularly in regard to the degree of stability or fragility, and none of them explicitly consider the reporting of social norms, all of which are important for external validity.

We recommend developing guidance on impact evaluation that takes into account study contexts at a fundamental level. Such guidance should consider the design of studies and not only the reporting of specific methods. It should also guide the choice of methods for constructing counterfactuals to suit study contexts. Lastly, it should

encourage flexibility and transparency in judgements about ensuring study designs are suited to their contexts.

## Section 1

### **Introduction**

The Centre of Excellence for Development Impact and Learning (CEDIL) develops and tests innovative methods for evaluation and evidence synthesis that can inform socio-economic development in low- and middle-income countries (LMICs). Innovation is required when methods that are generally accepted as rigorous cannot be employed in certain contexts, such as in rapid emergencies, or in relation to certain initiatives, like peace-building programmes.

This paper provides an overview of how, in CEDIL-funded studies, claims arising from research (termed 'evidence claims' for brevity) have been justified and communicated in order to inform policy decisions relating to socio-economic development. The CEDIL-funded studies analysed comprised three programmes of work: evaluating complex interventions; enhancing evidence transferability; and increasing evidence use. Phase I of the research explored how the CEDIL-funded studies that we analysed were designed to justify their claims,. Phase II analysed the final reports on the CEDIL-funded studies to assess how claims were justified in practice.

In this paper, we first present our research problem and questions (Section 2), and then we present the study design, including a description of the data available and the method use for the analysis (Section 3). We then present and discuss our first set of study findings in terms of the wider literature, before presenting a framework for developing, justifying and communicating evidence claims (Section 4). Finally, we apply this framework to the final reports of the CEDIL studies (Section 5), and draw our conclusions and make recommendations (Section 6).

## Section 2

### Research problem and questions

In cases where decisions are informed by evidence from research, increased confidence tends to be placed in evidence that results from systematic reviews of multiple studies that take into account the studies' quality. In systematic reviews, the validity, reliability and trustworthiness of study findings are often expressed in terms of the quality of study methods, the standards to which studies adhere, the methodological criteria studies meet, or studies' strengths and limitations. These study criteria relate to how well a study is designed and conducted: i.e. its *internal validity*. Decision makers are also particularly interested in how *transferable* study findings are to a specific setting, or how *generalisable* they are to any setting beyond the one in which the study was conducted. Criteria to support these judgements assess a study's *external validity*. When study authors provide clear justifications for internal and external validity, this help decision makers (and other end-users) place appropriate confidence in a study's findings and claims (i.e. their suggested answer to a given research question). Understanding these justifications may be easier when they are clearly communicated in standardised ways.

When we began this study, we anticipated that standardisation would be most common and clearest within systematic review networks, where applying criteria to assess study quality is an integral part of the work. However, a survey of 14 publicly available English-language evidence portals in Europe and USA found that evidence standards vary across portals, even though these portals focused predominantly on questions about the efficacy of interventions that are answered by conducting experimentally controlled studies (Gough and White, 2018). Some portals make evidence claims based on individual studies, while others rely on an evidence base that brings together what is known from many studies or from evidence-informed guidance for policy and practice. Portals vary on the level at which they make evidence claims, and on the sophistication of the evidence standards that are used at each level. Gough and White (2018) recommended that those developing evidence portals should 'specify and justify the different evidence standards for making different claims about the existing evidence base including impact, strength, extent and consistency of evidence, process and contexts and costs' (p40).

We anticipated that standardisation would be least common where methods are newly developed, such as in the CEDIL programme, a core aim of which is to develop new methods.

This raises two important questions:

- When research is designed to produce findings for decision makers, how are claims arising from the findings justified? In other words, what criteria (evidence standards) are study authors using to justify their evidence claims?
- How are claims and justifications communicated in original research reports and in other outputs that share the research more widely?



## Section 3

### Study design and methods

The overall purpose of this study is to identify the lessons learnt from a series of empirical studies (systematic reviews and primary studies) that were commissioned by CEDIL for the shared purpose of developing and testing innovative approaches to evaluating impact and synthesising evidence in low-income countries.

To achieve this objective, the study first compiled and analysed formal research outputs (CEDIL design papers or final study reports) and less formal communications (blogs and webinars). The analysis of formal and informal study outputs took into account information from CEDIL's guidance for study authors contributing to its research programme, and learning opportunities provided in the form of a workshop convened by CEDIL in March 2022 to discuss evidence to enable policy, practice and decision-making.

The analysis distinguished between different kinds of claims across three dimensions:

The first dimension distinguishes between claims resting on systematic reviews and claims resting on single primary studies. Claims (and their justification) resting on systematic reviews depend on both the data accumulated from the full set of studies included in a systematic review, and the methods applied in the systematic review. Claims (and their justification) resting on single primary studies may be apparent from primary study reports and from criteria used to appraise their quality when included in a systematic review.

The second dimension distinguishes between claims made about the studies in the given context (specific claims about what happened) and claims made about similar hypothetical studies in other contexts (generalisable claims about what is likely to happen elsewhere).

The third dimension distinguishes between (a) empirical claims based on findings arising from substantive questions, and (b) methodological claims based on findings arising from methodological questions. For instance, substantive questions may lead to claims about population movements following disasters, while methodological questions may lead to claims about how well mobile phone data can track population movements.

The study was conducted in two phases. Phase I, which was conducted early on in CEDIL's programme of commissioned studies, resulted in a framework for developing, justifying, and communicating evidence claims. Phase II, conducted as that programme of commissioned studies was coming to an end, applied that framework to better understand the challenges in, and advances made in, making and justifying claims, particularly about complex interventions and the transferability of evidence.

**Phase I** analysed data that were available early on in the programme, including the following: guidance from CEDIL; discussions during a workshop that was convened to discuss evidence claims; 15 design papers for empirical and methodological (but not conceptual) studies commissioned by CEDIL; and blogs. The recurrent themes relating

to empirical claims that were apparent in the CEDIL programme were seen to match an established framework for developing and evaluating complex interventions (Skivington *et al.* 2021, p. 1). Recurrent themes relating to methodological claims that were apparent in the CEDIL programme were developed into an analogous framework for developing and evaluating research methods.

**Phase II** analysed 12 of the 13 reports relating to studies that remained in the CEDIL programme after CEDIL support for some studies ended due to UK aid cuts during the COVID-19 pandemic (one study had not reported by then).

The final outputs of two studies were available during Phase I. Analysis of an evidence portal (Shakespeare <https://disabilityevidence.org>) contributed to developing the framework, and so is reported as part of the discussion of Phase I. The analysis of an evidence synthesis (Rathinam *et al.*, 2020b) contributed to understanding claims made about complex interventions, and therefore is reported as part of the discussion of Phase II.

## Data we analysed

### Guidance from CEDIL

CEDIL provided guidance (undated) to the teams it commissioned, on planning stakeholder engagement and facilitating evidence for decision-making 'by planning for engagement with policymakers, practitioners, research commissioners, evaluators and other key stakeholders throughout the life of the study' (p 1).<sup>1</sup> This is relevant to all CEDIL-commissioned teams when considering how to influence decision makers through their own work, and for the CEDIL studies commissioned in the programme of work about increasing evidence use.

The CEDIL guidance implies a link between the confidence that is placed in evidence and the uptake of the findings when it encourages teams to ask:

'Is evidence easily accessible to policymakers, programme managers, research commissioners and other stakeholders? Is there capacity to appraise the relevance and quality of evidence?'

Other references to 'claims' in this guidance relate to evidencing the use of research by capturing citations about the work, records of meetings or feedback about the work, mentions in policy or programme documents, decisions, funding, guidelines, organisational strategies, and the designing or commissioning of subsequent studies; as well as by engaging with implementing agencies about monitoring and evaluation systems or institutional structures.

### A workshop on evidence to enable policy, practice and decision-making

While the commissioned studies were still underway, CEDIL convened a workshop on 29 March 2022 to better equip research teams to influence policy and practice through having in place **a clear influence strategy** and **a credible basis on which to evidence claims**. Participants were invited to reflect on both their influence strategies and the basis of their evidence claims, and thus how these claims will have an influence on policy and practice. The study teams commissioned by CEDIL were invited to critically assess the claims they expected to make. To meet the objectives of the workshop (circulated with the agenda), participants were prompted to consider that:

'systematic reviews, critical appraisal methods are used to assess the confidence we should have in both individual study findings, and the summaries of bodies of evidence.<sup>2</sup> The confidence we have may vary according to the "ambition" of the claim, that is "this worked here" compared to "this will work everywhere".'

Thirteen study teams were invited to attend the virtual workshop on Zoom. Their studies took the following forms: evidence syntheses (7), evaluations (4), secondary data analysis (1), and an exploratory study (1).

---

<sup>1</sup> <https://cedilprogramme.org/wp-content/uploads/2020/03/CEDIL-stakeholder-engagement-and-evidence-use-plan-guidance-final.pdf>

<sup>2</sup> Collections of critical appraisal tools may be found at <https://www.cardiff.ac.uk/specialist-unit-for-review-evidence/resources/critical-appraisal-checklists> and <https://jbi.global/critical-appraisal-tools>

The opening presentation by CEDIL's director discussed 'evidence standards, evidence claims and policy influence' (White, 2022). Three study teams presented their work at the workshop; one of these (and a fourth team) submitted reflections in response to questions circulated after the workshop. Other participants also contributed to the workshop discussion. The discussion was recorded and was automatically transcribed. Written contributions to the Zoom platform chat were also recorded.

Following the workshop, the funders circulated a note to participants reminding them that one of the goals of the workshop was that the discussion would lead to a 'lessons learnt' paper on how the CEDIL-funded projects aim to influence policy. Those not able to attend were asked to send a short note (not longer than one page) answering the following questions:

- What is the approach of your research project to influence policy?
- What evidence claims will you make and what confidence can we have in these claims?
- How will the evidence inform policy, practice or decision-making?

### **Study documents**

The original proposal for the work reported in this paper focused on CEDIL-funded systematic reviews. The scope was broadened when CEDIL invited authors of both systematic reviews and primary studies to attend the aforementioned workshop. We therefore analysed the outputs from 15 studies (evidence syntheses or primary research studies), which addressed the following:

- strengthening the methods used to evaluate complex, multi-component interventions and to improve theoretical understanding of causal chains that explain how and why combinations of activities work (eight studies: four evidence syntheses and four evaluations);
- developing and testing middle-range theories to explain how programmes work in a plurality of contexts and how interventions can be designed and adapted to novel contexts (five studies: two syntheses, one evaluation, and two secondary data analyses); and
- assessing stakeholder engagement, and making sense of evidence and communication methods (two studies: both syntheses).

In addition to the presentations and reflections produced for the workshop, outputs from these studies included design papers, webinars and blogs.

### **CEDIL conference**

A CEDIL conference titled 'Innovations in Impact Evaluation: What Have we Learned?', lasting five and a half hours and spread over four days (Tuesday 21 to Friday 24 February 2023) was held as CEDIL's work was approaching to an end, after five years. The conference provided an opportunity to note recurrent learning themes (although there was insufficient time for in-depth analysis of presentations or discussions).

## Data analysis

### Identifying key issues and current debates from the workshop

We read the workshop presentations, transcript and chat in full and highlighted quotes that were relevant to the topic of evidence claims. Academic papers mentioned by participants informed the discussion in Section 5 of the present paper.

### Extracting and analysing data about claims from study outputs

We read the CEDIL guidance to identify the common approach that commissioned teams were encouraged to take. We inspected research outputs to identify the following: explicit and implicit criteria in research proposals; quality appraisal tools; and interim study outputs and final reports (including text reporting study methods and findings, and discussing the studies' strengths and limitations).

We adopted the following initial rapid methods to inspect study outputs: reading specific sections of reports (methods, discussion of strengths and limitations) and reading sections of text identified by searching study reports for terms that might identify key text: claim\*, justif\*, reliab\*, trustworth\*, quality, standard\*, criter\*. Text discussing results and conclusions were also read, for the few completed studies available.

After familiarisation with some of the study outputs, the following questions were formulated:

- What kinds of empirical claims are made, and how are they justified?
- What kinds of methodological claims are made, and how are they justified?
- Is the justification explicit or implicit?

Two authors worked independently to apply these questions to all of the design papers, and then they checked and added to each other's work.

We read in full, shorter, less formal, interim outputs, such as blogs or webinars for a wider readership, because it was considered that their lay language may involve the use of wider terms. We applied the following questions to these outputs:

- How is the evidence claim communicated for different audiences?
- Can confidence in the claim communicated be traced back to the original assessments?



## Section 4

### Phase I: Developing an evidence claims framework

Seven CEDIL studies published design papers (i.e. research protocols) after the CEDIL guidance (undated) became available but before the more in-depth consideration of evidence claims that took place in the workshop presentations and discussion (in March 2022).

Table 1 describes the aims of these design papers. Two studies focused on empirical advances: one relating to intervention impact (Rahman *et al.*, 2022) and another relating to the mechanisms underlying intervention impact (Chioda and Gertler, 2022). One study focused on the methodological goal of being able to make better inferences and predictions in regard to other contexts (Davey *et al.*, 2022). The others were explicit about both their empirical and methodological aims (Burchett, 2022; Maselko *et al.*, 2022; Abdulrahim, 2022; de Brauw, 2022).

Whether the design papers' aims were to advance empirical knowledge or methodology, all papers made (or implied) both empirical and methodological claims. Advances in empirical knowledge require claims about methodology. Advances in methodology rest on empirical claims. This is addressed in the next two sections.

We also identified other project outputs for wider communication, including webinars and blogs. These are considered in the sub-section on communicating evidence claims.

**Table 1: Empirical and methodological aims of design papers**

Authors Type of study	Title	Empirical aims	Methodological aims
Abdulrahim, 2022 Evaluation	Gender-Sensitive Risks and Options Assessment for Decision Making (ROAD) to Support WiF2	Apply a 'complete ROAD process by focusing on decent work for women through safe migration pathways. ROAD is an overarching framework and facilitated learning process that starts with identifying risks associated with a challenge.' (p. 5)	Develop, test and validate the Women's Empowerment in Migration Index (WEMI). (p. 9)

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Authors Type of study	Title	Empirical aims	Methodological aims
Burchett, 2022 Evidence synthesis	Upstream interventions aiming to encourage adolescents' use of contraception in low- and middle-income countries: A rationale and protocol for a mixed-methods synthesis to develop a mid-range theory	(1) to develop a mid-range theory to explain how upstream interventions can encourage adolescents' use of contraception in LMICs. Within this, the authors aimed to explore what types of interventions have been evaluated, what intervention characteristics may facilitate or hinder their effectiveness, and what mid-range theory could explain how these interventions achieve effectiveness. (p. 1)	'to reflect on how best to build a mid-range theory using novel methods within an evidence synthesis.' (p. 1)
Chioda and Gertler, 2022 Evaluation	Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (8-year follow-up)	To address 'key policy-relevant knowledge gaps identified by the literature (1) improving the understanding of which skills—and which combination of skills—are important for leadership and entrepreneurship and how to teach them; (2) the sustainability of these interventions' impacts; (3) documenting any spillovers beyond the usual economic outcomes, such as risky behavior and IPV; (4) identifying the mechanisms at play underlying effects as well as the subgroups that are most likely to benefit from these types of programs.' (p. 2)	

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Authors Type of study	Title	Empirical aims	Methodological aims
Davey <i>et al.</i> 2022 Evaluation, secondary data analysis	POInT Research Design Paper		'The goal of our project is to develop and implement a flexible and accessible method for combining data from the impact evaluation and data from the process evaluation to make better inferences about the theory of change and the transportability of the results as predictions for other contexts.' (p. 3)
de Brauw, 2022 Evaluation	Impact Evaluation of the SHARPE Project in Ethiopia	SHARPE was initially designed to improve refugee livelihoods and self-reliance, and generate economic opportunities for host communities through the piloting and scaling of market systems development programmes across seven market systems sectors: crops, livestock, fish, wood, labour, aid, and access to finance. In designing market systems development programmes, along with economic actors already present in the systems being studied, SHARPE aims to create interventions that both strengthen the economies of refugee-hosting populations and reduce tensions between refugee and host populations.	The aim is to combine the following: (a) the inherent adaptability of long-term market systems projects as they continually monitor changes in market systems for further constraints on growth or development; with (b) the pre-determined characteristics of short-term experimental designs to evaluate effectiveness – by focusing the latter on key constraints to uptake for rapid evaluation.

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

<b>Authors</b> <b>Type of study</b>	<b>Title</b>	<b>Empirical aims</b>	<b>Methodological aims</b>
Maselko <i>et al.</i> , 2022  Evaluation	Impact of Maternal Depression Treatment on Maternal Health, Parental Investment, and Child Development	'(1) Evaluate the impact and provide a cost-effectiveness analysis of the randomized intervention, the Thinking Healthy Programme Peer-delivered (henceforth THPP) on maternal health and child development through age 6 years. (2) Identify behavioural and biological mechanisms through which the THPP influenced maternal health and child development. (3) Assess the impact of providing personalized feedback to parents regarding their child's performance on future parenting behaviours.' (p. 1)	'(1) Going beyond the estimation of average treatment effects, we plan to use novel machine learning techniques to estimate heterogeneity of treatment effects and identify baseline characteristics of women that are predictive of their responsiveness to treatment.' (p. 1)
Rahman <i>et al.</i> , 2022  Evidence synthesis	Technology-Based Innovative Solutions for Improving Perinatal Care Utilization: A Network Meta-Analysis	'This systematic review and network meta-analysis aim to fill this gap in knowledge by synthesizing the impact of technology-based interventions on a comprehensive range of reported outcomes in ANC, PNC, and delivery care. Furthermore, our meta-analysis will evaluate the impact by sociodemographic, country and regional factors, which will further inform policymakers on the potential use of technology-based interventions in given contexts.' (p. 4)	

## Phase I: Current debates highlighted by the workshop

### Setting the scope of the discussion

The opening presentation (White, 2022) warned against making claims based on studies that are too general, and assuming that conclusions from a single study can be widely applied, and recommended instead basing claims on systematic reviews of all available evidence (as such reviews more commonly assess the strength of evidence claims).

White set the scope of the discussion about evidence standards, evidence claims and policy influence in terms of the following:

- What the claims are about – such as the effectiveness or relevance of an intervention (e.g. the prevalence of the problem it addresses), technical feasibility and acceptability (in a formative evaluation), implementation issues (in a process evaluation), and cost-effectiveness analysis.
- What evidence claims are drawn from (e.g. an individual study, summaries of bodies of evidence, or research and broader implications) and transferred to (e.g. guidance documents).
- The standards and tools used to assess the confidence we can have in study findings.

White explained that common standards and tools for assessing quality are largely accepted for evidence synthesis, but are less so for primary studies. There is greater consensus about standards and tools for quantitative research than those for qualitative research. While the authors of individual studies may justify the quality of their work, many users of evidence do not read the original studies (or reviews). Instead, they access summaries of evidence through portals that present evidence in simplified forms. These portals apply standard criteria when presenting conclusions about impact, and confidence in that impact. Thus, having in place common evidence standards is particularly important for people who do not read the original reviews. A widely used example is the Teaching and Learning Toolkit<sup>3</sup> produced by the Education Endowment Foundation, which is widely used by teachers.

Following this introduction, study teams were invited to consider their evidence claims in terms of what sorts of claims they were and the degree of confidence they had in their claims.

### Reflections from CEDIL study teams

Presentations at the workshop emphasised engaging stakeholders and contextual differences over evidence standards. There was the typical portrayal of 'strong evidence' from RCTs and 'weaker evidence' from mixed-methods studies. Following the workshop, reflections about evidence claims raised issues about confidence in claims based on mixed methods (presentation 1), communicating the impact and transferability of evidence (presentation 2), and methodological advances (presentation 3).

---

<sup>3</sup> <https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit>



*Presentation 1.* A framework synthesis on violence outcomes of water, sanitation and hygiene (WASH) interventions led to evidence claims about the risk of violence being linked to sanitation infrastructure, social acceptance of violence and lack of consequent punishment, and the wider built environment, and about social cohesion and group protection facilitating feelings of safety (Macura *et al.*, 2022). As this broad range of issues was identified from studies that applied various methods, confidence in the claims came from applying a mixed-methods critical appraisal tool (MMAT) to identify and synthesise only those studies that were of high or medium quality.

*Presentation 2.* According to the presentation, a secondary data analysis of four RCTs that were ongoing at the time (Davey *et al.*, 2022) was expected to claim that ‘process and impact data can be combined using a formally modelled version of the theory of change, and that this allows for insights that are not possible to discern otherwise, as well as other efficiency gains’. The team anticipated having moderate confidence in the claim after successfully applying the method in at least three case studies, for the first time bringing together various sources of information about an intervention using the theory of change as a key component of the analysis. According to the presentation, it was considered that additional research would be required before the method could be widely and easily adopted by other teams. In particular, it was said that ‘using qualitative data to inform the parameters of a causal model is very difficult given the current standards for conducting qualitative research. This is because qualitative inquiry about causal processes rarely investigates all of the possible contextual factors that might affect whether a causal process is thought to take place’.

*Presentation 3.* The presentation explained that an evidence synthesis of education<sup>4</sup> calculated evidence of impact from random-effects meta-analysis. The authors stated, ‘[headline claims are] presented next to a padlock rating that describes the extent and quality of the evidence. For example, the meta-analytic effect of delivering feedback to pupils is  $d=0.481$ . This is communicated as 6 months progress and give a padlock rating of 4 out of 5, with the approach losing a padlock due to the number of studies that do not have randomised allocation between groups.’ As explained in the presentation, this visual communication relies on applying standardised criteria to the evidence. Added to this (according to the presentation) is local evidence focusing on feasibility and implementation, often qualitative evidence, with no claims made about impact (Okwen, 2022).

In summary, these three presentations illustrated the following: applying and justifying evidence standards that are well-established; increasing the confidence we can have in evidence claims by applying methods to multiple case studies with a shared theory of change; and communicating confidence in evidence claims visually.

---

<sup>4</sup> <https://cedilprogramme.org/funded-projects/programme-of-work-2/using-meta-analysis-to-explore-the-transferability-of-education/>

### Reflections from workshop discussants

Three discussants considered the policymakers' perspective on using evidence, the academic perspective on the use of methodological innovations, and the link between confidence in evidence claims and influencing policy.

#### **Box 1: Discussants at the workshop**

**James Copestake** is Professor of International Development at the University of Bath. His research ranges across agrarian change and rural development, development finance and its evaluation, conceptualisations of poverty and wellbeing, and the political economy of development and development studies. He is Co-director of the Centre of Development Studies, Director of Studies for the professional doctorate in policy research and practice at the Institute for Policy Research, and a founding director of Bath SDR Ltd, a social enterprise dedicated to improving qualitative and mixed method impact evaluation.

**Ian Goldman** has worked for 40 years on issues of rural development, decentralisation, sustainable livelihoods approaches, community-driven development, and evidence-based policymaking and implementation, applying the disciplines of planning, evaluation, research and action-learning. He has worked in 20 countries in Africa, Europe and Latin-America; with national, provincial and local governments; and in the non-governmental organisation and small and medium-sized enterprise sectors. From 2011 to June 2018 he was Head of Evaluation and Research in South Africa's Department of Planning, Monitoring and Evaluation (DPME), where he led the establishment of South Africa's national evaluation system, and in 2017 became Deputy Director General for the new Evaluation, Evidence and Knowledge Systems Branch. In 2015 Ian was the founding Chair of the Management Committee of the Twende Mbele African M&E Partnership, with partners Uganda, Benin, South Africa, CLEAR AA and African Development Bank, and he has been a Commissioner of the Initiative for Impact Evaluation (3ie) since 2012. Ian left DPME in July 2018 to join the Centre for Learning on Evaluation and Results (CLEAR Anglophone Africa), based at the University of Witwatersrand in Johannesburg, advising internationally on evaluation and evidence, teaching and writing, and he is leading a research project leading to a book on evidence-based policy in Africa. In addition, Ian is an Adjunct Professor at the Nelson Mandela School of Public Governance at the University of Cape Town, Visiting Professor at the University of Reading in the UK, and Honorary Research Associate at South Africa's Human Sciences Research Council.

**Sandy Oliver** is Professor of Public Policy at the UCL Institute of Education. For 30 years her interests have focused on the interaction between researchers and people making decisions in their professional and personal lives, largely through conducting systematic reviews. She is a member of the Board of the Campbell Collaboration and a Cochrane editor with their Consumers and Communication Review Group. She works with the UK Foreign, Commonwealth and Development Office (FCDO) to build capacity in systematic reviewing in developing countries.

Ian Goldman took a policymaker's perspective. He emphasised the importance of the relationship between those producing evidence and those who are well-placed to use it. He signposted an evidence use framework developed in Africa. This framework centres the demand for evidence, which then influences both the nature of the evidence subsequently generated and its potential for use in decision-making (Langer and Weyrauch, 2020). Elements to consider about the generation of evidence include: 'the quality of the supplied evidence (e.g. is it trustworthy and relevant); the type of the evidence (e.g. are research questions and methods fit-for-purpose to address the policy need); and the specific evidence claim (e.g. does the body of available evidence and existing evidence standards support the recommendation)' (Langer, 2021, p. 3) and the quality of the entire process from prior to generation (e.g. developing terms of reference), through generation and the use of interventions to encourage use which triggers change mechanisms of awareness, agreement, access, interaction/trust, ability, institutionalisation/formalisation. Goldman shared South African government evidence standards for evaluation.<sup>5</sup> According to Goldman, the focus of these is demand-led rather than researcher-led, and they cover quantitative and qualitative evaluation, not just synthesis. As Goldman expressed in the workshop, the standards encompass:

'the planning and design and adequacy of resourcing the appropriate design... the implementation, the evaluation ethics, participation and skills development, methodological integrity, the project management and then the reporting... all rated and scores used when... presenting these evaluations to cabinet'.

James Copestake encouraged interaction with professional evaluation associations to promote the uptake of methodological innovations. He also emphasised the value of wider peer review to promote 'being more explicit about how synthesis and interpretation takes place.' This includes helping to locate systematic reviews as part of a wider set of responses to the challenge of how to generalise, including realist synthesis and sensemaking workshops. 'Standards' and 'use' combine here to the extent that programming and budgeting open-ended post-analysis deliberative interpretation and contextualisation of findings can be important to both.

Sandy Oliver focused on the link between the two themes in this workshop: confidence in evidence claims and influencing policy. She suggested that there is a loose connection between the two, with some claims being more influential even if confidence is low, and vice versa. This is inevitable when policymakers take into account other factors beyond the research, such as ease of change or political expediency. However, the strength of the connection may be influenced by the research itself. For instance, confidence may be lower for evidence claims about sub-groups, specific settings, or intervention providers, yet the importance of these factors may sway decisions. Indeed, these factors may be particularly influential even if confidence in the evidence claims is, from a researcher's perspective, too low. This raises questions about whether engaging decision makers at the stage of research design can strengthen confidence in the

---

<sup>5</sup> Available at

<https://www.dpme.gov.za/keyfocusareas/evaluationsSite/Evaluations/DPME%20Standards%20for%20Evaluation%20in%20Government%20v2%2014%2003%2006.pdf>

evidence claims that matter most to them, and what the trade-offs are between evidence that is generalisable and evidence that is context-specific.

In summary, the workshop discussion emphasised the following: starting with the demand for evidence in order 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 engaging in wider peer review to ensure greater clarity of synthesis and interpretation.

## Phase I: Making and justifying claims in CEDIL design papers

Our analysis of the CEDIL design papers identified empirical claims and revealed how these claims were justified. The claims and their justifications aligned closely with an existing framework for developing and evaluating complex interventions. The methodological claims made in these design papers were aligned with an analogous framework for developing and evaluating research methods. The details are presented in the subsections below.

### Empirical claims in study design papers

The empirical claims we identified in the CEDIL design papers, and their justifications, are listed in Appendix 3. These claims included claims about **gaps in prior knowledge** and the **importance of those gaps** and **research questions**, and claims about what **knowledge studies were expected to advance**.

Claims about knowledge gaps identified gaps in the existing research literature. For instance, the claim that '[no studies] assess the complex interplay and impact of technology-based interventions over the dynamic antenatal-delivery-postnatal pathway' was supported by citing related (but different) studies, although the absence of a systematic review was not explicitly stated (Rahman *et al.*, 2022, p. 6). Similarly, the justification that the study would fill a gap was implicit in the choice of methods and tools. Chioda and Gertler (2022) justified their claim that there is a knowledge gap by referring to the emerging findings of an ongoing systematic review about business training programmes, and justified its importance by citing the findings from employer surveys. Davey *et al.* (2022) claimed to go a step further, by helping evaluation teams formulate the right 'high-impact evaluation questions', although how this was done, other than reflecting on existing evidence, was unclear (no page numbers).

More detailed claims were made about the **effects of interventions, key elements for interventions and for implementation**, intervention **constraints**, local **relevance**, and **usefulness**.

In regard to claims about assessing the effects of interventions, justification was implied by the choice and details of established methods (Abdulrahim, 2022; Maselko *et al.*, 2022; Nakamura, Leyew *et al.*, 2022; Handa, 2022). In these cases, readers were left to judge the appropriateness of the methods. Where comparison groups were established in different ways (randomisation, discontinuity design using the proxy means test cut-off, and matched wards (administrative units below the district), followed by the application of household targeting by the programme so all households in the

comparison wards are future eligible households), readers were required to judge the implications of these different designs (Handa, 2022).

A claim regarding generating an effect size with greater confidence relied on innovative methods, and was justified by argument (Davey *et al.*, 2022). A claim regarding identifying key elements for interventions and implementation was explicitly justified by the choice of, and detailed explanations of, recently developed methods (qualitative comparative analysis (QCA) and intervention component analysis (ICA)) (Burchett, 2022, p. 4). However, the quality appraisal methods chosen for possibly excluding low-quality studies were not defined in advance.

Recognising that the intervention effects may be constrained, de Brauw . (2022) claimed their study would identify constraining factors, address them with intervention amendments, and evaluate those amendments using rapid RCTs. This empirical claim was supported by examples of constraints already identified by qualitative research (complemented by quantitative work) to understand the support factors, derailers and intermediate steps between intervention and desired outcomes operating in the study contexts.

Two design papers made claims about enhancing the local relevance of their work through engaging stakeholders. In one paper, it was explained that the opportunity for data to be collected by 'female research assistants with robust training who thoughtfully attend to the cultural context of the area' came from strong working relationships developed over 20 years (Maselko *et al.*, 2022, p. 4). There was an implicit assumption that this particular method of data collection would yield 'better' findings. Burchett (2022) held a similar assumption about engaging stakeholders in advisory groups to shape the development of the mid-range theory. This was partially justified by citing other authors. However, this design paper acknowledge that 'there is much left to be understood around how the involvement of stakeholders can clarify the concepts represented, whether additional adverse impacts are accounted for with the input of stakeholders, how the involvement should be managed, and what happens when there are divergent views between the stakeholders and the evidence' (Burchett, 2022, p. 7). No solutions were offered in the description of methods.

Another study (Macura *et al.*, 2022) emphasised the relevance of findings in a blog post, statement: 'The review scope is co-designed with stakeholders ensuring relevance of findings for WASH policy and practice.'<sup>6</sup> This statement is justified by the fact that feedback on the protocol was made publicly available.<sup>7</sup>

Chioda and Gertler (2022) claimed their findings would be useful to policymakers because they 'identify which set of skills and/or program components are effective levers such that, if activated by other programs, they could move outcomes along the causal chain' (p. 23). The justification for this claim was implicit in the choice of methods;

---

<sup>6</sup> <https://www.sei.org/projects-and-tools/projects/advancing-evaluation-of-gender-equality-outcomes-in-wash/>

<sup>7</sup> <https://www.sei.org/wp-content/uploads/2020/05/response-to-feedback-from-open-consultation-1.pdf>

readers need to understand the methods in order to be able to assess their potential benefits.

### Framework for developing and evaluating complex interventions

Taken together, these empirical claims can be seen to rest on well-established methods for developing and evaluating complex interventions. Early efforts to standardise such methods resulted in a framework that focused on evaluating complex interventions with RCTs (Campbell *et al.*, 2000). Later developments accommodated additional formative evaluations, process evaluations, and observational designs in cases where experimental designs cannot be applied, and tailoring interventions to local circumstances (Craig *et al.*, 2008). Figure 1 presents the current framework for methods that are ‘commonly used in the health and social care services, public health practice, and other areas of social and economic policy that have consequences for health’ (Skivington *et al.*, 2021, p1). Placing the context of evaluation and engaging stakeholders at the core of the latest framework, 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.

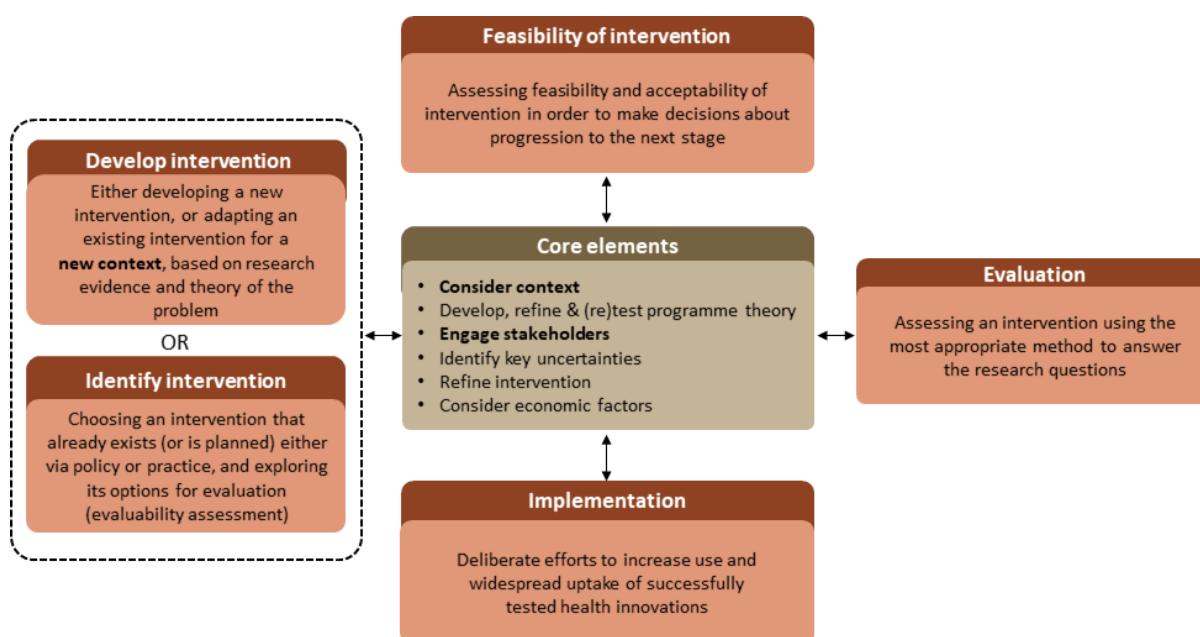


Figure 1: Framework for developing and evaluating complex interventions (Skivington *et al.*, 2021)

### Methodological claims made in study design papers

Methodological claims and their justifications are listed in Appendix 4. Methodological claims in design papers or reports included claims about **choice of methods**, **innovative methods for measuring effect sizes**, **confidence in causal claims**, **taking research findings from one context and using them in another**, and **innovative combinations of methods**.



One study, completed in Phase I, about using big data relied on a systematic map of existing studies<sup>8</sup> as a proof of concept for big data methods contributing to the evidence base in development sectors where evaluations are often infeasible due to data issues (Rathinam *et al.*, 2020a; 2020b). The study also noted caveats. In particular, very few studies report on big data ethical issues: informed consent, data privacy, data security and unintended exclusion. Moreover, several analytical, ethical and logistical challenges may hinder the use of big data in evaluations. The study called for standards to be set for the reporting of data quality issues, data representativeness and data transparency.

For studies in progress, where design papers employed well-established methods, the justification was implicit: readers were assumed to understand the strengths of case studies (Davey *et al.*, 2022), network meta-analysis (Rahman, 2022), machine learning (Maselko *et al.*, 2022), qualitative research and mixed methods (Abdulrahim, 2022).

The choice of recently developed methods for developing mid-range theory with QCA and ICA and stakeholder engagement was justified in more detail, with citations to prior research (Burchett, 2022).

Similarly, Handa (2022) aimed to adopt 'recent developments at the intersection of machine learning (ML) and causal inference to build a "middle-range" theory around poverty traps... indicating which households are more likely to benefit from a cash transfer versus other types of interventions, and what the pathways are to sustained increases in consumption' (pp. 6–7). This claim was justified explicitly by a detailed argument and by citing earlier methodological studies.

Innovative methods for measuring effect sizes, increasing confidence in causal claims or targeting transportability (Davey *et al.*, 2022) were justified by argument, and by referencing the high quality of the case studies that provided the data. Similarly, Chioda and Gertler (2022) claimed all their measures had been validated and tested, and cited papers validating instruments in the context of interest (p. 12).

Two design papers claimed innovations in the form of interdisciplinary working. Davey *et al.* (2022) provided 'a protocol for integrating not just quantitative and qualitative data, but also the theoretical insights of different practitioners' (p. 35) This is achieved by grounding team discussions in diagrams with boxes and arrows that depict assumptions about underlying causal processes. This approach was justified by argument, and by citing prior authors' work on boundary objects for interdisciplinary working. Maselko *et al.* (2022) employed innovative combinations of impact measures and study designs. They combined biomarkers with 'an unusually rich set of measures of child development, maternal function and parental investments, using state of the art psychometric tools' uniquely embedded in an RCT (p. 7). While each design and method was well-established, and was described with supporting citations, the unique combination was justified by argument.

The SHARPE evaluation also combined different research paradigms: market systems projects that were inherently adaptable, and RCTs which were researcher-controlled (de

---

<sup>8</sup> <https://gapmaps.3ieimpact.org/evidence-maps/big-data-systematic-map>



Brau, 2022). The empirical claim these approaches support is mentioned above. This same claim can also be interpreted as a methodological claim about appropriate combinations of research methods. The feasibility of adding rapid RCTs to long-term projects to inform subsequent intervention adaptations was justified by carefully constructing an initial theory of change and applying primarily qualitative methods to identify intervention constraints which might offer opportunities for intervention improvements. In summary, claims resting on well-established methodologies tended to be implicit. In contrast, claims of methodological innovation were supported more explicitly by detailed argument, with 'proof of concept' studies or by citing other methodological studies. Where innovative methodological claims rested on argument alone, empirical data supporting these arguments may subsequently appear in the final reports.

### Framework for developing and evaluating research methods

Taken together, these methodological claims can be seen to rest on well-established approaches for developing and evaluating research and evaluation methods. They can be brought together in a framework (Figure 2) that is analogous to the framework for developing and evaluating complex interventions (Figure 1).

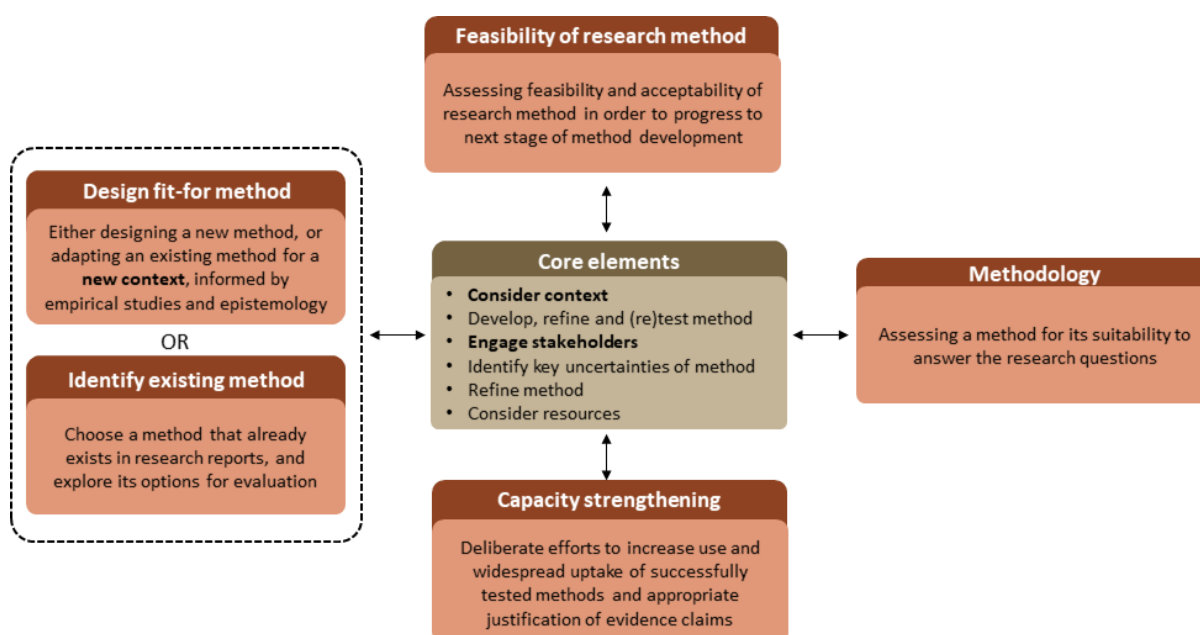


Figure 2: Framework for developing and evaluating research methods

### Phase I: Communicating claims for other audiences

Most ongoing CEDIL studies with design papers in the public domain had other outputs in the public domain that were linked to from their home page on the CEDIL website: blogs, a policy brief, a working paper, and a presentation (see Appendix 2).

The working paper (Chioda *et al.*, 2012) adopted largely the same methodological claims and justifications as the design paper, mentioning the RCT study design, but only included discussion about one of the interventions. It adopted largely the same technical language as the design paper. This is appropriate for two outputs targeted at the same type of audience and not designated for the general public.

The policy note (Wu *et al.* 2021) used similar language as its associated design paper to communicate the issue of female migration and forced labour, and provided a similar depth of explanations. While not directly referencing the study, the policy recommendation to the G20 also made reference to Sustainable Development Goal 8.7, just like the design paper, to justify the study's relevance and importance in a global context.

The Disability Evidence Portal<sup>9</sup> offers public access to evidence, in the form of policy briefs, to inform disability-inclusive development. It uses a visual scale to communicate how feasible, accurate, relevant and inclusive the research is that is synthesised to develop each policy brief. How this has been done is justified and communicated in a publicly available Rating System Appraisal Checklist (Appendix 6).

Some studies had associated blogs. One described the focus and methods of the study (Burchett, 2022). It adopted an interview format and plain language, for an expert to communicate the rationale of the project to a general audience. While going into less technical detail, the blog post was much more explicit when justifying methodological claims about the appropriateness of the methods to answer the research question and the benefits of the tools chosen for analysis, namely QCA and ICA.

The second blog (Sikander, 2021) discussed how COVID-19 affected field research in general, and specifically addressing this study. Unlike its technical design paper (Maselko *et al.*, 2022), it was written in accessible language, avoiding jargon, explained concepts and implications in detail, and highlighted the potential limitations imposed by COVID-19.

The third blog was published before its design paper was available (Macura *et al.*, 22).<sup>10</sup> Therefore, the claims about stakeholder involvement ensuring relevance, and its justification with publicly available feedback on the protocol,<sup>11</sup> were clearly communicated to a wide readership.

The SEEDS study of the long-term impacts of a complex agricultural intervention on welfare, behaviour and stability in Syria justified empirical claims by providing a blog<sup>12</sup> which provided direct access to visual data in the form of Google Earth historical imagery, which showed that more recent photographs were less green. Similarly, the methodological claims were partially justified by providing interactive images. This blog

---

<sup>9</sup> <https://www.disabilityevidence.org/index.php/questions-evidence>

<sup>10</sup> <https://www.sei.org/projects-and-tools/projects/advancing-evaluation-of-gender-equality-outcomes-in-wash/>

<sup>11</sup> <https://www.sei.org/wp-content/uploads/2020/05/response-to-feedback-from-open-consultation-1.pdf>

<sup>12</sup> <https://storymaps.arcgis.com/stories/1392deed227f45a98e2093c922a03798>

provided links to underpinning publications, although these are not necessarily funded by CEDIL.

One completed study developed an evidence portal that makes explicit claims of relevance to people with disabilities in LMICs (Shakespeare <https://disabilityevidence.org>). These claims, and how they were justified, were clear from a tool made publicly available that assessed the evidence in terms of stakeholder involvement, local relevance, and feasibility in variety of resource settings.

Clearly communicating evidence claims to wider audiences, visually and through blogs, may motivate and facilitate the use of evidence for decision-making.

## Phase I: Discussion

### Phase I: Summary of findings

**Nature of claims:** By focusing on the analysis of evidence claims in a set of studies from the Global South, we have broadened discussion about the nature of claims in several dimensions. The original emphasis of evidence claims on the effects of intervention in areas where synthesised evidence is plentiful has broadened to include methodological claims as well as empirical claims, claims resting on single studies and bodies of evidence, claims about developing and implementing interventions (not only their effects), and 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, 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 were 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 these claims is not always communicated clearly: authors may rely on readers understanding 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.

### **Phase I: Strengths and limitations**

This paper draws on the experience and insights of academics and policymakers with an impressive track record of research spanning a broad range of academic disciplines and methodologies, with links to policy decision-making in various contexts. These academics and policymakers offered insights during the workshop discussion and signposted relevant academic papers and policy briefs. Nevertheless, conducting methodological debates across such a broad scope is challenging and the CEDIL programme itself offers a very small number of studies. Unfortunately, David Gough, whose scholarship has advanced thinking about evidence claims (Gough, 2021), was unable to join the workshop or lead this paper because of illness.

Research for socio-economic development is very varied, with single studies sometimes cutting across academic disciplines, research methodologies and policy sectors where discussion of criteria or standards for assessing evidence claims is uneven. The broad scope of the discussion workshop therefore presented challenges to individual participants, whose familiarity with the debates within the scope of the discussion was necessarily limited by their specific experience. For instance, the authors of this paper are familiar with some but not all of the methodologies adopted by the CEDIL studies. Contributions to the workshop would probably have been more developed if the first presentation had been available earlier, allowing for more preparatory time to consider the broad range of issues that evidence claims cover, what they are based on, and the availability (or lack in some areas) of standards and tools for assessing the confidence we can have in study findings.

Only six CEDIL design papers were available for our analysis. We found other outputs for ongoing studies, but not for studies no longer supported by CEDIL as a result of UK aid cuts during COVID-19. We did not have time to listen to webinars. Our interpretations of the CEDIL documents and workshop discussion have not been validated by the study authors or workshop participants.

In summary, the CEDIL studies and workshop provided a fruitful avenue to discuss and explore evidence claims and their justification in socio-economic development. Ongoing discussions will be strengthened by drawing on wider literature.

### **Phase I: Related literature**

This paper has drawn together from CEDIL studies examples of evidence claims and their justifications about a broad range of issues. These examples involve judgements that are largely instrumental (about the nature of the claim and whether it is relevant to

users' needs) and technical (employing standards, tools and guides for appraising the evidence). However, Gough's (2021) high-level framework for appraising the 'fitness for purpose' of evidence claims includes two other components:

- the values underpinning claims and whether other claims, arising from questions with different framings, need to be considered too; and
- engaging with the evidence, taking into account other information, such as information about opportunity costs.

The second of these components was raised by discussants in the workshop. Copestake noted how deliberative interpretation and contextualisation of findings can be important for both standards and their use. Oliver surmised that confidence in evidence claims may depend on the standpoint (such as a policy or research perspective) of those making the judgement. These social aspects of evidence claims reframes them from the technical sources from which they are drawn (e.g. an individual study, summaries of bodies of evidence, research and broader implications, or guidance based on study findings) to who is making the claims and how. Claims may be made by authors or reviewers of research (either individual studies or summaries of bodies of evidence) with or without taking into account broader implications. While the technical aspects of making claims can be standardised by applying tools to assess the confidence we can have in study findings (see Figure 1 in Gough, 2021), the social aspects of such decisions are influenced by the numbers, background and hierarchy of those involved, the time and knowledge resources available, and the degree of formality in seeking consensus (Oliver *et al.*, 2015). Recognising standards in these social aspects began in the health sector with a tool for assessing the quality and reporting of practice guidelines. This tool assesses the quality of the evidence available and the range of stakeholders involved (Brouwers *et al.*, 2010) but not how the group deliberates. The evaluation literature includes similarly broad conceptions for justifying claims, using the language of validity: for the preparatory work when values are first introduced, the technical design and evaluation conduct, reporting, and interpretation and use beyond the evaluation team (Downes and Gullickson, 2022). In this literature, there is a gap in the knowledge about evaluation criteria and standards relating to negotiating evaluation purposes and questions.

The CEDIL studies that were analysed sometimes justified their claims with argument, rather than evidence of effects, particularly when using novel methods. This distinction may be analogous to the distinction Melendez-Torres *et al.* (2016 115) make between 'practical-configurational modes of reasoning and inferential-predictive modes of reasoning' employed in narrative and meta-analytic reviews, respectively. These authors call for specific quality criteria for reviews employing practical-configurational reasoning to making sense of the evidence.

Some CEDIL studies made claims about the effects of interventions. Rahman (2022) justified these claims by reference to the use of accepted methods and tools for assessing study quality. These were tools for assessing the internal validity of studies. Given the importance of the relevance of evidence to decisions being made, there is also a role for methods and standards for assessing external validity or transferability:

in other words, the extent to which the outcomes of a successful intervention evaluated in a primary context can be achieved in a target context. Using this definition, Schloemer and Schröder-Bäck (2018) systematically reviewed the health literature and then developed two models: a conceptual model that explains the underlying mechanisms of transferability, and a process model that guides its assessment.

Munthe-Kaas *et al.* (2020) complemented this work by focusing on issues related to context during the systematic reviewing process. After discussions with stakeholders about current practices and needs, they systematically identified transferability checklists and developed a seven-stage process with guidance and templates relating to taking transferability into account throughout the planning and conducting of systematic reviews.

When tools for assessing studies included in a review are time-consuming to apply, they may not align with policymakers' timetables for decisions. In these circumstances, researchers may streamline tools to ensure faster evidence delivery (see Chrisp *et al.* (2022) for an example). Such an approach raises questions about quality standards for context-specific evidence that meets the needs of specific decision makers, and quality standards for evidence that can justifiably be used repeatedly by different decision makers.

CEDIL study teams who made evidence available for decision makers presented visually appealing summaries with simple scales or images to communicate their confidence in evidence claims. Because some decisions require considering options or complex interventions, multidimensional images have been developed to present research evidence from complex evidence syntheses with multiple outcomes, Cochrane reviews, and clinical guidelines (Babatunde *et al.*, 2018).

### **Phase I output: Framework for developing, justifying and communicating evidence claims**

Empirical claims made by CEDIL-funded research teams align with the framework for designing and evaluating complex interventions (Figure 1, Skivington *et al.* (2021)), and the methodological claims align with an analogous framework for developing and evaluating research methods (Figure 2). The emphasis on contextual considerations and stakeholder engagement at the core of these frameworks particularly suits settings that are resource-poor or unstable, such as LMICs, humanitarian settings and areas affected by conflict. However, they remain research-led rather than fitted to the purpose of decision-making. Neither framework considers issues raised in the CEDIL workshop: interventions' or methods' relevance, importance, and 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 we developed a framework (Figure 3) that emphasises the following: the demand for evidence (as recommended in the CEDIL workshop) so as to design research that strengthens confidence in the evidence claims that matter most for decision-making; and the values and perspectives of decision makers, not only researchers (as recommended by Gough (2021)). The core tasks begin with articulating



evidence needs and engaging stakeholders. The framework balances the conceptualisation of evidence claims with the processes of engaging stakeholders. Indeed, the two are necessarily intertwined as stakeholders debate and justify claims, systematically appraising them both conceptually and technically, with support from guidance and tools relating to both technical tasks and the collaborative process.

The concepts that contribute to the ‘nature of the claims’ and the ‘stakeholder engagement’ in our new framework (Figure 3) could form the starting point for analysing evidence claims and their justification in empirical and methodological studies.

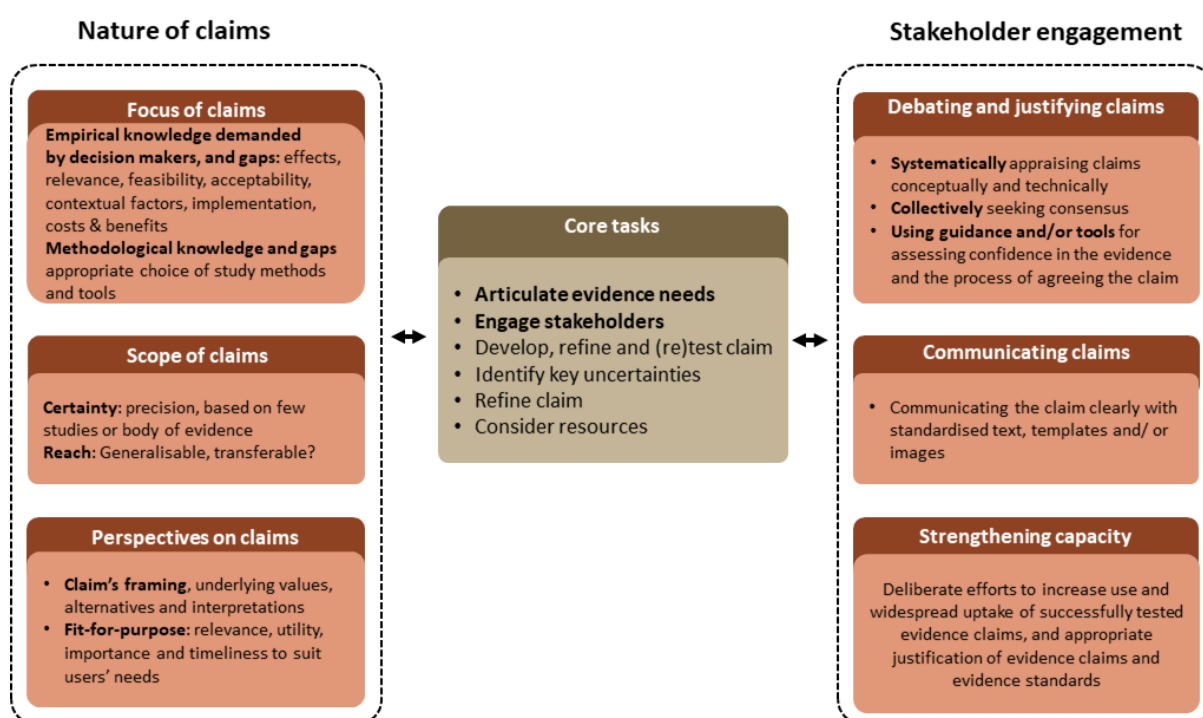


Figure 3: Framework for developing, justifying and communicating claims



## Section 5

### Phase II: Applying an evidence claims framework

The framework developed in Phase I is applied here to the final reports of CEDIL studies of complex interventions to examine the claims made, and their justifications, about:

- evidence for developing effective complex interventions (their focus);
- the claims' certainty and reach (their scope); and
- the claims' framing and fitness for purpose (their backdrop).

#### Phase II: Focus of claims about developing effective complex interventions

The CEDIL studies that were analysed made claims about complex interventions by investigating how they work, the effects seen at different levels (households and communities), and the contextual factors that mediate or moderate the effects (sometimes considered in terms of equity).

Claims regarding the **effectiveness of complex interventions** were supported by standard methods of statistical meta-analysis. For instance, quantitative analysis using r with robust variance estimation, meta-analysis and meta-regression supported the claim that:

'The most effective interventions [for involving men and boys in family planning] are community-based educational programmes offered in schools, communities and homes or community facilities, and interventions involving multiple components, delivered by professionals, trained facilitators or peers to both males and females for over seven months. Brief programmes of less than three months are also effective.' (Aventin *et al.*, 2023)

Similarly, statistical meta-analysis with subgroup analyses identified differential effects to draw conclusions about **equity**:

'there are substantial time savings from water supply interventions, which largely accrue to women, as well as substantial time savings from sanitation accruing to both women and men.' (Macura *et al.*, 2022)

However, studies available for synthesis rarely provided the data required for extended analyses of equity to encompass a specific caste or class, the elderly, ethnicity, religion, or people with disabilities; and never provided data on gender and sexual minorities (Macura *et al.*, 2022).

In one evaluation, claims about equity over the longer term relied on a methodological innovation that applied machine learning to a combination of panel data about households and their socio-economic position and satellite data about the agricultural productivity of communities (Baliki *et al.*, 2023). This study found that 'female-headed households benefit remarkably more in terms of food security in the medium-term compared to male-headed families' (p. 3).

Some studies extended these methods to consider **how complex interventions work**, with what key components or in what circumstances. Although the language differed, three studies employed similar methods, the essential steps being to propose the chain of events that an intervention might trigger (a logic model or theory of change), then testing that proposal empirically with a causal chain analysis (Aventin *et al.*, 2023), or by aligning quantitative and qualitative evidence along the theory of change (Nakamura, Molotsky *et al.*, 2022). This is how evidence was developed to support empirical claims:

'The success of family planning programmes that involve men and boys is most often measured by contraceptive use to the relative neglect of other outcomes, such as met need for family planning, equitable family planning decision-making, or gender equality.' (Aventin *et al.*, 2023)

'the main activities necessary for a successful [language of instruction] program are the presence of high-quality teaching and learning materials in the [mother tongue]... Intermediate outcomes of Mother Tongue Education (MTE) included positive perceptions of results related to teaching quality and increased student motivation while final outcomes from MTE studies include perceived improvements in first language and second language reading skills, especially in comparison to students who were not studying in the MTE schools.' (Nakamura, Molotsky *et al.*, 2022).

A third study that began by constructing a theory was a secondary analysis of evaluation data. It tackled complexity by translating that theory into a detailed flow diagram – a directed acyclic graph (or DAG) – which 'precisely and succinctly represents a large number of causal claims which add up to a complicated theory of intervention causation' (Juden *et al.*, 2023, p. 14). These causal claims were then tested by applying Bayesian causal modelling to help evaluators make better inferences about the theory of change and the transportability of the results to other contexts.

Two studies unpicked complexity from the other direction. In an evidence synthesis, Burchett *et al.* (2022) analysed studies using QCA and ICA in order to develop a mid-range theory for tailoring interventions to the specific life stage of an adolescent, with activities to enhance motivation, agency and access to contraception.

In an evaluation of national unconditional cash transfer programmes, Handa *et al.* (2022) also started with the data, using new machine learning algorithms to build a theory of change by starting with identifying those households most likely to benefit from unconditional cash transfer programmes – an automated method of positive deviance inquiry, which helps make sense of heterogeneity in complex contexts. They saw younger households with fewer dependents diversifying their livelihoods, making a productive shift towards the market, and thereby raising their standard of living significantly after four years.

For two studies, the theory of change at the centre of the analysis was developed with input from stakeholders. The scope of a review about WASH, with an emphasis on gender and social equity, was co-designed with stakeholders, ensuring the relevance of the findings for WASH policy and practice (Macura *et al.*, 2022). A logic model for involving men and boys in family planning, based on both stakeholder consultation and

prior research, provided a framework for synthesising qualitative research (Aventin *et al.*, 2023).

## Phase II: Scope of claims, in terms of certainty and reach

Our framework considers the scope of claims in terms of their certainty and their reach or transferability. While many CEDIL studies employed methods to maximise or assess the degree of certainty of their claims, only one put this at the centre of their research question.

Bernard *et al.* (2023) claimed that observational methods for assessing the effects of intervention have, on average, 'zero bias, but the bias distribution has high variance such that any given observational study is likely to have high bias.' They justified this claim by reanalysing RCTs in development economics that had incomplete compliance, to create hypothetical studies that compare either (a) the effects of offering an intervention (an intention-to-treat analysis) or (b) the effects of receiving an intervention. This claim is discussed again in the sub-section below about the backdrop for claims, in terms of their framing and fitness for purpose.

More studies made claims about the transferability of evidence, which raises questions about how such claims are justified and with what success. Many of these were the same studies that made claims about developing effective interventions by clarifying and testing **theories of change**. In their study of involving men and boys in family planning, Angrist and Meager (2022) combined studies of effectiveness and implementation to identify influential contextual factors, such as gender norms and how they interact with structural factors such as education and labour markets.

Similarly, Aventin *et al.* (2023) not only made claims about suitable settings, intervention components and delivery (see the sub-section above on the focus of claims): conclusions from their meta-analysis and subsequent RCT also highlighted the importance of fidelity when delivering a programme in a new context. The mid-range theory developed by Handa *et al.* (2022) using machine learning identified two contextual features associated with households performing better in agriculture: rainfall and (in one country) living closer to the district capital, with less land devoted to agriculture.

A study of complex agricultural interventions acknowledged the potential impact of environmental shocks, such as hyperinflation, drought, violent conflict, and COVID-19, although without detailing exactly their influence in a theory of change (Baliki *et al.*, 2023). Nevertheless, this study illustrated the importance of taking into account context, finding that 'households residing in areas that are moderately affected by violent conflict show stronger food security improvements [following complex agricultural interventions] compared to households from peaceful or conflict-intense settings' (p 7).

Some studies took into account context by noting differences across countries. For instance, in their synthesis, Macura *et al.* (2022) noted that most studies were conducted in particular geographical regions, and raised the possibility of variations in social and cultural dynamics leading to different gender outcomes elsewhere. Higgins *et al.* (2022) explored differences in norms in more depth. The aim was to re-contextualise

an evidence portal from the English education system to suit Cameroon, Chad, Nigeria and Niger. Judgements about the relevance for these African countries were made by local policymakers and practitioners. They considered that most pedagogical approaches with evidence accessible through the portal were largely appropriate and applicable in their own contexts. However, five approaches (one-to-one tuition, school uniforms, setting and streaming, teaching assistants' interventions, and learning styles) that had evidence supporting their use in the UK they considered either too expensive or impractical for other reasons in the local contexts. Interventions related to another three topics (menstrual hygiene interventions, cash transfers, and corporal punishment) were considered particularly useful: the first two because such interventions could encourage school attendance (especially by girls), and the last because evidence could help discourage current practices.

## **Phase II: Backdrop for claims, in terms of their framing and fitness for purpose**

The last feature in our framework about the nature of claims is their backdrop, in terms of the values underlying them and the purpose for which they are generated or may be used.

The backdrop for CEDIL's programme of work is the relative lack of evidence suiting the Global South, particularly for **areas requiring humanitarian aid** where **decisions are urgent, access is difficult and situations are unstable**. The backdrop in terms of research **methodology** is **impact evaluation and evidence synthesis** for decision-making.

The underlying values of impact evaluation are the assumptions about strengthening claims about causal relationships, with an emphasis on randomisation. However, among the CEDIL studies that were analysed, an impact evaluation in Ethiopia confirmed problems with access, and problems with mounting RCTs in areas of conflict (de Brauw 2023 et al). Empirical claims were limited. The promise of conducting rapid RCTs was an illusion, with the time required to manage them much longer than the time allowed to deliver research findings. As things turned out, local staff losses following political conflict, and COVID-19 travel restrictions, affected the research team's ability to build a local understanding of the context, to work on the relationships needed to manage the project, and to solve problems from a local perspective. This contrasts with the study mentioned above (Bernard *et al.*, 2023) that, after comparing RCTs with hypothetical observational studies, concluded by recommending randomisation because the results offer less variance. In this case, only one of the RCTs included was mounted in a conflict-affected area.

Another study conducted in areas affected by conflict avoided those problems: in part, by using the novel data collection method of satellite photography, and by applying alternative methods for constructing counterfactuals. It adopted adaptive research designs with flexible matching and balancing to improve the comparability of groups for rigorous impact evaluation (Baliki *et al.*, 2023).

The backdrop for many of the analysed studies, although this was not necessarily visible, was the inequality embedded in gender norms. Gender norms were given a high profile in the evidence syntheses about family planning (Aventin *et al.*, 2023) and contraception (Burchett, 2022), but were hardly mentioned in other studies, even when noting livelihood interventions that had differential effects for female-headed households (de Brauw, 2023). The exception was a synthesis of the gender and social outcomes of WASH interventions, particularly as regards gender-based violence (Macura *et al.*, 2022). However, a key observation of this study was that most studies about WASH interventions do not address the potential transformational impacts for girls and women, such as eliminating violence against women and girls, education, women's economic empowerment, and women's participation and leadership in WASH services.

## **Phase II: Discussion**

### **Phase II: Summary of findings**

In the CEDIL study reports, comparing empirical data with theories of change was central to both identifying active components and intervention pathways, and to identifying contextual factors that may influence the intervention effects. Involving stakeholders is valuable for both developing theories of change, and for understanding important contextual influences. Advances in machine learning similarly help make sense of vast amounts of complex data for both tasks. While prospective impact evaluations, with randomisation, in theory offer strong evidence for attributing causality, such studies may be impractical in areas affected by conflict or where evidence is needed urgently. Largely missing from the CEDIL study reports was explicit attention to social norms to frame studies, and to the multidimensional aspects of diversity and inequity for analysing their data or interpreting their findings.

### **Phase II: Strengths and limitations**

A key strength for this phase of the project was having a framework to support the analysis. This framework is based on (a) the shared knowledge of academics and policymakers with a specialist interest in evidence claims for informing policy decisions, and (b) a growing methodological literature about impact evaluation and evidence synthesis, including stakeholder engagement and the transferability of findings. A key limitation was the small number of studies available for analysis.

### **Phase II: Wider methodological literature**

The strengths and limitations of the studies analysed can be seen as largely reflecting the consensus or standards in the methodological literature. This becomes apparent if one searches the extensive set of standards that have been collated by the EQUATOR (Enhancing the QUALity and Transparency Of health Research) Network,<sup>13</sup> an international initiative that seeks to improve the reliability and value of published health research literature by promoting transparent and accurate reporting and wider use of robust reporting guidelines.

---

<sup>13</sup> <https://www.equator-network.org/>

Many of the reviews analysed above adopted quality standards, such as recording how studies were identified with a PRISMA flow diagram – although not necessarily recording alignment with other PRISMA reporting standards (Page *et al.*, 2021) or the guidelines for reporting equity (Welch *et al.* 2012). Similarly, the RCTs adopted accepted conventions for assessing causal relationships, without necessarily reporting alignment with international standards.

Conversely, the EQUATOR Network does not appear to offer standards for alternative methods for constructing counterfactuals that are more suitable for evaluating impact in areas affected by conflict. Nor does the reporting of social norms appear in EQUATOR listed guidelines.

However, the EQUATOR Network does signpost a checklist, based on a systematic review, for reporting theories of change, although this is focused on theories of change for developing and evaluating public health interventions (Breuer *et al.*, 2016). This checklist covers why and how a theory of change has been developed, who was involved, and how the resulting theory was used to develop interventions. Although not considering socio-economic interventions more broadly than in the public health domain, some of the included studies included in the systematic review were conducted in LMICs.

A more recent systematic rapid review addresses how evidence has been integrated into theory of change frameworks in the healthcare sector by drawing on institutional data, the research literature and key stakeholders (Romão *et al.*, 2023). These authors note the limited overlap between the literature about theories of change and evidence-informed decision-making. Evidence syntheses that employ theories of change, as seen in several CEDIL studies, may provide a bridge between these two literatures.



## Section 6

### **Summary, conclusions and recommendations**

From discussions at the CEDIL workshop, outputs from CEDIL design papers and related literature, we constructed a framework for developing, justifying and communicating claims about empirical research findings and research methods. These claims can vary in terms of:

- their focus: aspects of evidence for developing effective complex interventions;
- their scope: the claims' certainty and reach or transferability; and
- their backdrop: the claims' framing and fitness for purpose.

The nature of such claims results from engaging stakeholders with each other to debate and justify claims, preferably working systematically and collectively, using guidance and/or tools to assess the technical aspects and the collaborative aspects of the work. Subsequent steps for enhancing evidence-informed decision-making include communicating claims clearly, and strengthening capacity in both research and engagement skills for developing and justifying evidence claims.

Applying this framework to the CEDIL study reports confirms that opaque, changing and unpredictable contexts present major challenges to longitudinal prospective studies (typically RCTs) in regard to making claims attributing causality. In the study reports, such challenges were overcome in part by drawing on the knowledge of diverse stakeholders to understand the contexts where they live and work, and to develop theories about how interventions may play out in particular contexts. Other partial solutions included collecting data remotely (for instance, via satellite or mobile phones), analysing vast amounts of data with machine learning or artificial intelligence, and conducting multidimensional analysis of population diversity and equity. Although methods are advancing rapidly in these two areas, many of the studies paid scant attention to social norms or the complexities of diversity.

Widely accepted standards encourage evaluations that demonstrate internal validity. Such standards were not necessarily apparent in the CEDIL study reports we analysed. Moreover, the standards themselves pay scant attention to the reporting of study contexts, particularly the degree of stability or fragility, and none of them explicitly consider the reporting of social norms.

We recommend guidance being developed for impact evaluation that takes into account study contexts at a fundamental level. Such guidance should consider the design or framing of studies, not only the reporting of specific methods. It should also guide the choice of methods for constructing counterfactuals to suit study contexts, and the development and use of theories of change for policy sectors beyond health. Lastly, it should encourage flexibility and transparency in judgements about suiting study designs to their contexts.



## References

### CEDIL Programme references

- Abdulrahim, S. (2022) 'Gender-Sensitive Risks and Options Assessment for Decision Making (ROAD) to Support WiF2', *CEDIL Design Paper 4*, CEDIL, London and Oxford.
- Angrist, N. and Meager, R. (2022) 'The role of implementation in generalisability: A synthesis of evidence on targeted educational instruction and a new randomised trial', *CEDIL Syntheses Working Paper 4*, CEDIL, Oxford. Available at: <https://doi.org/10.51744/CSWP4>
- Aventin, Á., Robinson, M., Hanratty, J., Keenan, C., Hamilton, J., McAteer, E. R., Tomlinson *et al.* (2023) 'Involving men and boys in family planning: A systematic review of the effective components and characteristics of complex interventions in low- and middle-income countries', *Campbell Systematic Reviews* 19, e1296. Available at: <https://doi.org/10.1002/cl2.1296>
- Balasubramanya, S., Stifel, D., Alvi, M., Ringler, C. (2022) 'The role of social identity in improving access to water, sanitation and hygiene (WASH) and health services: Evidence from Nepal', *Dev Policy Rev.* 40, e12588. Available at: <https://doi.org/10.1111/dpr.12588>
- Baliki, G., Brück, T., and Ghattas, H., (2021). "Syria: Do Agricultural Interventions Strengthen Food Security?" CEDIL Evidence Brief 1, *CEDIL, Oxford*. Available at: <https://doi.org/10.51744/CEB1>
- Baliki, G., Weiffen, D., Al Daccache, M., Kayaoglu, A., Sujud, L., Jaafar, H., Ghattas, H., & T. Brück (2023). Seeds for recovery: The long-term impacts of a complex agricultural intervention on welfare, behaviour and stability in Syria (SEEDS), CEDIL Research Project Paper 7. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford. Available at: <https://doi.org/10.51744/CRPP7>
- Bernard, D.R., Bryan, G., Chabe-Ferret, S., de Quidt, J., Fliegner, J., Rathelot, R. (2023) 'How biased are observational methods in practice? accumulating evidence using randomised controlled trials with imperfect compliance', CEDIL final academic report, CEDIL Research Project Paper 9, *CEDIL, Oxford*. Available at: <https://doi.org/10.51744/CRPP9>
- Burchett, H. E. D. (2022) 'Upstream interventions aiming to encourage adolescents' use of contraception in low- and middle-income countries: A rationale and protocol for a mixed-methods synthesis to develop a mid-range theory', *CEDIL Design Paper 1*, CEDIL, London and Oxford.
- Burchett, H., Kneale, D., Griffin, S., Joaquim Picardo, J., de Melo, M. and French, R. (2022). Structural interventions aiming to enable adolescent use of contraception in low- and middle-income countries, CEDIL Syntheses Working Paper 5, CEDIL, Oxford. Available at: <https://doi.org/10.51744/CSWP5>
- Chioda, L., Gertler, P. (2022) 'Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training

- programs in Uganda (8-year follow-up)', *CEDIL Research Design Paper 8*, CEDIL, London and Oxford.
- Chioda, L., Contreras-Loya, D., Gertler, P., Carney, D. (2021) 'Making Entrepreneurs: Returns to Training Youth in Hard Versus Soft Business Skills', *CEGA Working Paper series No WPS-156*, Centre for Effective Global Action. University of California, Berkeley.
- Davey, C., Juden, M., Allan, E., Prost, A., Humphreys, M., Jacobs, A., Medina, L., Mwambi, H. (2022) 'POInT Research Design Paper', *CEDIL Research Design Paper 3*, CEDIL, London and Oxford.
- de Brauw, A. (2022) 'Impact Evaluation of Ethiopia's SHARPE program', CEDIL Design Paper 7. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford.
- de Brauw, A., D. Gilligan, L., Leavens, F., Moges, S., Roy, Tefera M. (2023) 'Impact Evaluation of the SHARPE Programme in Ethiopia: Academic Report', CEDIL Research Project Paper 6, CEDIL, London and Oxford. Available at: <https://doi.org/10.51744/CRPP6>
- El Didi, H., C. van Biljon, M. Alvi, C. Ringler, N. Ratna, S. Abdulrahim, P. Kilby, J. *et al.* (2022) 'Reducing vulnerability to forced labour and trafficking of women migrant workers from South- to West-Asia', *Development in Practice. Special issue on Modern Slavery and Exploitative Work Regimes*. Available at: <https://doi.org/10.1080/09614524.2022.2059448>
- Handa, S. (2022) 'An Empirically-driven Theory of Poverty Reduction', CEDIL Design Paper 6. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford.
- Handa, S., Liu, Z., Tembo, G., Adamba, C. & Mvula, P. (2023) 'An empirically driven theory of poverty reduction', CEDIL Research Project Paper 4. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford. Available at: <https://doi.org/10.51744/CRPP4>
- Higgins, S., Dobson, E., Kay, J., Okwen, P., Kamga, E., Nsaikila, M., Hanny, R. *et al.* (2022) 'Re-contextualising an evidence portal from the English education system to Cameroon, Chad, Nigeria and Niger', *CEDIL Research Project Paper 2*, CEDIL, London and Oxford. Available at: <https://doi.org/10.51744/CRPP2>
- Juden, M., Mapuwei, T., Tietz, T., Sarguta, R., Medina, L., Prost, A., Humphreys, M *et al.* (2023) 'Process Outcome Integration with Theory (POInT)', academic report, CEDIL. CEDIL Research Project Paper 5. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford. Available at: <https://doi.org/10.51744/CRPP5>
- Langer, L., Weyrauch, V. (2020) 'Using evidence in Africa', in 'Using Evidence in Policy and Practice: Lessons from Africa', Routledge, London.
- Maselko, J., Baranov, V., Biroli, P., Bhalotra, S., Hagaman, A., Bates, L.M., Turner, E.L., Sikander, S. (2022) 'Impact of Maternal Depression Treatment on Maternal Health,

- Parental Investment, and Child Development', *CEDIL Design Paper 2*, CEDIL, London and Oxford.
- Macura, B., Dickin, S., Sharma Waddington, H., Liera, C., Soto, A., Orlando, A. *et al.* (2022) 'Gender and social outcomes of WASH interventions: synthesis of research evidence', *CEDIL Syntheses Working Paper 7*, CEDIL, London and Oxford. Available at: <https://doi.org/10.51744/CSWP7>
- Nakamura, P., Leyew, Z., Molotsky, A., Ranjit, V. and Kamto, K. (2022) 'Language of Instruction in Schools in LMICs: A Systematic Review', *CEDIL Design Paper 12*, CEDIL, London and Oxford.
- Nakamura, P., Molotsky, A., Castro Zarzur, R., Ranjit, V., Haddad, Y., de Hoop, T. (2023) 'Language of instruction in schools in LMICs: a systematic review', *Campbell Systematic Reviews*, *in press*
- Okwen, P. (2022) 'Recontextualising the Teaching and Learning Toolkit to the Chad Basin For CEDIL', presentation to FCDO on 14 February 2022.
- Oliver *et al.* (2021) 'Engaging Stakeholders with Evidence and Uncertainty: Developing a toolkit', *CEDIL Methods Working Paper 4*, CEDIL, London and Oxford. Available at: <https://doi.org/10.51744/CMWP4>
- Peterson, C., te Vogt, S. & Lal Fak Mawii, J. (2022) 'Enhancing evaluation use: An experimental assessment of rubric methodology on collective judgements in Myanmar', *CEDIL Design Paper 11*. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford.
- Rahman, M.O., Islam, M.R., Rahman, M.S., Taniguchi, H., Jung, J., Lohan, M., Ota, E., Rahman, M.M. (2022) 'Technology-Based Innovative Solutions for Improving Perinatal Care Utilization: A Network Meta-Analysis', *CEDIL Design Paper 5*, CEDIL, London and Oxford.
- Rathinam, F., Khatua, S., Siddiqui, Z., Malik, M., Duggal, P., Watson, S., Vollenweider, X. (2020a) 'Using big data for evaluating development outcomes: a systematic map', [Online], 3ie. Available at: <https://gapmaps.3ieimpact.org/evidence-maps/big-data-systematic-map>
- Rathinam, F., Khatua, S., Siddiqui, Z., Malik, M., Duggal, P., Watson, S, and Vollenweider, X. (2020b) 'Using big data for evaluating development outcomes: a systematic map', *CEDIL Methods Working Paper*, CEDIL, Oxford.
- Rathinam, F., Khatua, S., Siddiqui, Z., Malik, M., Duggal, P., Watson, S, and Vollenweider, X. 2020. Using big data for evaluating development outcomes: a systematic map. *CEDIL Methods Working Paper 2*. Oxford: Centre of Excellence for Development Impact and Learning (CEDIL). Available at: <https://doi.org/10.51744/CMWP2>
- Sikander, S. (2021) 'COVID in Pakistan: its effects on field research work and future implications', *Bachpan cohort*. Available at: <https://www.bachpanstudy.com/post/covid-in-pakistan-its-effects-on-field-research-work-and-future-implications>

White, H. (2022) 'Evidence standards, evidence claims and policy influence' (with content from David Gough), 29 March 2021.

Wu, J., Kilby, P., Abdulrahim, S., Choudhury, Z.A., Ratna, N., El Didi, H., Ringler, C. (2021) 'Reducing vulnerability and precarity of low-skilled women in short-term migration from the global south: key policy recommendations for the g-20', policy brief. Available at: <https://www.t20italy.org/2021/09/22/reducing-vulnerability-and-precarity-of-low-skilled-women-in-short-term-migration-from-the-global-south-key-policy-recommendations-for-the-g-20/>

Wu, J., Kilby, P. (2022) 'The precarity of gender, migration, and locations: case studies from Bangladesh and Nepal', *Development in Practice*. Available at: <https://doi.org/10.1080/09614524.2022.2057441>

### Other references

Babatunde, O.O., Tan, V., Jordan, J.L., Dziedzic, K., Chew-Graham, C.A., Jinks, C., Protheroe, J., van der Windt, D.A. (2018) 'Evidence flowers: An innovative, visual method of presenting "best evidence" summaries to health professional and lay audiences', *Res Syn Meth.* 9, pp. 273–284.

Best, A., & Holmes, B. (2010). Systems thinking, knowledge and action: towards better models and methods, *Evidence & Policy*, 6(2), 145-159. Available at: <https://doi.org/10.1332/174426410X502284>

Breuer, E., Lee, L., De Silva, M., Lund, C. (2016) 'Using theory of change to design and evaluate public health interventions: a systematic review', *Implement Sci.* 11, p. 63. Available at: <https://doi.org/10.1186/s13012-016-0422-6>

Brouwers, M., Kho, M.E., Browman, G.P., Burgers, J.S., Cluzeau, F., Feder, G., Fervers, B. *et al.* (2010) 'AGREE II: Advancing guideline development, reporting and evaluation in healthcare', *Can Med Assoc J.* Available at: <https://doi.org/10.1503/cmaj.090449>

Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D. *et al.* (2000) 'Framework for design and evaluation of complex interventions to improve health', *BMJ* 321, p. 694. <https://doi.org/10.1136/bmj.321.7262.694>

Chrisp, J., Smyth, L., Stansfield, C., Pearce, N., France, R. and Taylor, C. (2022) 'Basic income experiments in OECD countries: A rapid evidence review', EPPI Centre, UCL Social Research Institute, University College London, London.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., Petticrew, M. *et al.* (2008) 'Developing and evaluating complex interventions: the new Medical Research Council guidance', *BMJ* 337, p. a1655. Available at: <https://doi.org/10.1136/bmj.a1655>

Downes, J., Gullickson, A.M. (2022) 'What does it mean for an evaluation to be 'valid'? A critical synthesis of evaluation literature', *Evaluation and Program Planning* 91. Available at: <https://doi.org/10.1016/j.evalprogplan.2022.102056>

Gough, D. (2021) 'Appraising Evidence Claims', *Review of Research in Education* 45, pp. 1–26. Available at: <https://doi.org/10.3102/0091732X20985072>

- Gough, D., White, H. (2018) 'Evidence standards and evidence claims in web based research portals', Centre for Homelessness Impact, London. Available at: [www.homelessnessimpact.org](http://www.homelessnessimpact.org)
- Hyder A., Syed S., Puvanachandra P., Bloom G., Sundaram S., Mahmood S., Iqbal M, Hongwen Z., Ravichandran N., Oladepo ., Pariyo G., Peters D. (2010) Stakeholder analysis for health research: Case studies from low- and middle-income countries. *Public Health* 124: 159-166. Available at: <https://doi.org/10.1016/j.puhe.2009.12.006>
- Langer, L. (2021) 'Using evidence in Africa: a framework to assess what works, how, and why', policy brief, CLEAR-AA, University of Witwatersrand, Johannesburg.
- Lapping K, Marsh DR, Rosenbaum J, Swedberg E, Sternin J, Sternin M, Schroeder DG. The positive deviance approach: challenges and opportunities for the future. *Food Nutr Bull.* 2002 Dec;23(4 Suppl):130-7. PMID: 12503241.
- Melendez-Torres GJ, O'Mara-Eves A, Thomas J, Brunton G, Caird J, Petticrew M. Interpretive analysis of 85 systematic reviews suggests that narrative syntheses and meta-analyses are incommensurate in argumentation. *Res Synth Methods.* 2017 Mar;8(1):109-118. Available at: <https://doi.org/10.1002/jrsm.1231>. Epub 2016 Nov 17. PMID: 27860329; PMCID: PMC5347877.
- Munthe-Kaas, H., Nøkleby, H., Lewin, S. *et al.* (2020) 'The TRANSFER Approach for assessing the transferability of systematic review findings', *BMC Med Res Methodol* 20, p. 11. Available at: <https://doi.org/10.1186/s12874-019-0834-5>
- Oliver, S., Hollingworth, K., Briner, R. (2015). 'Effectiveness and efficiency of committee work: a rapid systematic review for NICE by its Research Support Unit', Report of the Research Support Unit for the National Institute for Health and Care Excellence. EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London, London.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L. *et al.* (2021) 'The PRISMA 2020 statement: An updated guideline for reporting systematic reviews', *PLoS Med.* 18(3), p. e1003583. PMID: [33780438](https://pubmed.ncbi.nlm.nih.gov/33780438/)
- Romão, D.M.M., Setti, C., Arruda, L.H.M., de Melo, R.C., de Araujo, B.C., Tan, A.R. *et al.* (2023) 'Integration of evidence into Theory of Change frameworks in the healthcare sector: A rapid systematic review', *PLoS ONE* 18(3), p. e0282808. Available at: <https://doi.org/10.1371/journal.pone.0282808>
- Rychetnik L., Hawe P., Waters E., Barratt A., Frommer M. (2004) A glossary for evidence based public health. *J Epidemiol Community Health*; 58:538–545. Available at: <https://doi.org/10.1136/jech.2003.011585>
- Schloemer, T., Schröder-Bäck, P. (2018) 'Criteria for evaluating transferability of health interventions: a systematic review and thematic synthesis', *Implement Sci.* 13(1), p. 88. Available at: <https://doi.org/10.1186/s13012-018-0751-8>. PMID: 29941011; PMCID: PMC6019740.
- Skivington, K., Matthews, L., Simpson, S.A., Craig, P., Baird, J., Blazeby, J.M. *et al.* (2021) 'A new framework for developing and evaluating complex interventions: update of

Medical Research Council guidance', *BMJ* 374 , p. n2061. Available at:  
<https://doi.org/10.1136/bmj.n2061>

Tritter JQ, McCallum A. (2006) The snakes and ladders of user involvement: Moving beyond Arnstein. *Health Policy*. 76(2):156-68. Available at:  
<https://doi.org/10.1016/j.healthpol.2005.05.008>. Epub 2005 Jul 11. PMID: 16006004.

Welch, V., Petticrew, M., Tugwell, P., Moher, D., O'Neill, J., Waters, E., White, H, PRISMA-Equity Bellagio group (2012) 'PRISMA-Equity 2012 Extension: Reporting Guidelines for Systematic Reviews with a Focus on Health Equity', *PLoS Med.* 9(10), p. e1001333.

## **Acknowledgements**

---

We are grateful to James Copestake, Esther Forgan, Ian Goldman, Biljana Macura and Howard White, who commented on earlier drafts of this paper, and to participants at a workshop for FCDO, who discussed our work as it was being completed.



## Appendix 1

### Invitation to workshop

#### **CEDIL thematic workshop on evidence to enable policy, practice and decision-making, 29 March 2022**

#### **Objectives**

The purpose of the workshop is to better equip the research teams funded by CEDIL to influence policy. There are two aspects to this: having a clear influence strategy, and having a credible basis for evidence claims. Participants will reflect on both their influence strategies and the basis of their evidence claim, and thus how these claims will have an influence on policy (broadly defined). The presentations and the ensuing discussion will contribute to a CEDIL 'lessons learned' paper on promoting evidence use through evaluation and synthesis.

Link to virtual session:

<https://lshtm.zoom.us/j/91525169851?pwd=UXpKdjBqSVUyNURPT0JTV2hPL0I4UT09>

Meeting ID: 915 2516 9851

Password: 559663

Find your local number: <https://lshtm.zoom.us/u/aez85fjDTz>

<b>Allocated time</b>	<b>Session</b>	<b>Speaker and session content</b>
13.55–14.00	Arrival	Participants to join webinar session for a prompt start
14:00–14:15	Welcome and overview of event	CEDIL research directorate
14:15–14:30	Influencing strategies	Howard White – the session will discuss how the results of impact evaluations and evidence syntheses can influence policy
14:30–14:45	Evidence claims and policy impact	David Gough – the session will discuss how evidence claims can be credibly made and in what way they can be expected to affect policy
14:45–14:50	BREAK	

14:50–16:05	Speed presentations by up to 12 research teams	Each team will present for 5 mins, on the following: <ul style="list-style-type: none"> <li>• title of the study and brief description of methods and goals;</li> <li>• need, potential, and strategy employed, to influence policy;</li> <li>• the evidence claims to be made and the basis for being able to make these claims; and</li> <li>• how, in practice, the evidence will inform policy, practice or decision-making.</li> </ul>
16:05–16:15	BREAK	
16:15–17:00	Discussion and wrap-up	Plenary: questions, discussion and next steps

### CEDIL projects invited to the workshop (13)

Project name	Research team	Country	Programme of work (PoW)	Study type
'Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (7-year follow-up)'	Berkeley and World Bank	Uganda, Pakistan	PoW 2: Enhancing evidence transferability	Evaluation
'Putting the theory of change to work: Process-Outcome Integration with Theory'	London School of Hygiene and Tropical Medicine	Global	PoW 2: Enhancing evidence transferability	Secondary data analysis
'Seeds for recovery: The long-term impacts of a complex agricultural intervention on welfare, behaviour and stability in Syria (SEEDS)'	International Security and Development gCmbH (ISDC)	Syria	PoW 1: Evaluating complex interventions	Evaluation
'Gender-Sensitive Risks and Options Assessment for Decision-Making (ROAD) to Support WIF2'	International Food Policy Research Institute	Bangladesh, Jordan, Lebanon	PoW 1: Evaluating complex interventions	Evaluation

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

'Impact Evaluation of the SHARPE Project in Ethiopia'	International Food Policy Research Institute	Ethiopia	PoW 1: Evaluating complex interventions	Evaluation
'Using meta-analysis to explore the transferability of education mid-range theories to Cameroon, Chad, Nigeria and Niger'	Education Endowment Foundation	Cameroon, Chad, Nigeria, Niger	PoW 2: Enhancing evidence transferability	Evidence synthesis
'Scaling Social Accountability for Health: Leveraging Public Policies and Programs'	World Vision UK	Cambodia, Democratic Republic of Congo	PoW 2: Enhancing evidence transferability	Exploratory
'Gender and social outcomes of WASH interventions: synthesis of research evidence'	The Stockholm Environment Institute	Global	PoW 1: Evaluating complex interventions	Evidence synthesis
'A mixed-methods synthesis to develop a mid-range theory for interventions aiming to generate demand for contraception among adolescents'	London School of Hygiene and Tropical Medicine	Global	PoW 2: Enhancing evidence transferability	Evidence synthesis
'Understanding factors that influence Teaching at the Right Level's effectiveness and generalizability: A Bayesian Evidence Synthesis'	Young 1 ove Organisation	Global	PoW 2: Enhancing evidence transferability	Evidence synthesis
'Language Transitioning Research Synthesis'	American Institutes for Research	Global	PoW 2: Enhancing evidence transferability	Evidence synthesis
'How Biased are Observational Methods in Practice? Accumulating Evidence Using Randomised Controlled Trials with Imperfect Compliance'	University of Warwick	Afghanistan, Kenya, Nepal	PoW 3: Increasing evidence use	Evidence synthesis
'Involving Men and Boys in Family Planning: A systematic review of complex interventions to identify	Queen's University Belfast	Global	PoW 1: Evaluating	Evidence synthesis

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

effective programme components and characteristics in low- and middle-income countries'			complex interventions	
---	--	--	-----------------------	--

## Appendix 2

### Outputs from CEDIL's three programmes of work

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
<b>Programme of work 1: Evaluating complex interventions</b>							
'Big data for evaluating development outcomes (systematic gap map)' (Rathinam)	International Initiative for Impact Evaluation (3ie)	Multi-country	Study completed Phase I <ul style="list-style-type: none"> <li>Rathinam <i>et al.</i> (2020a). 'Using big data for evaluating development outcomes: a systematic map' [Online]. 3ie. Available <a href="#">here</a>.</li> <li>CEDIL <a href="#">Methods Working Paper</a> (Rathinam <i>et al.</i>, 2020b)</li> <li>CEDIL <a href="#">Methods Brief (Rathinam et al., 2021)</a></li> </ul>	Evidence synthesis			
'Seeds for recovery: The long-term impacts of a complex agricultural intervention on welfare, behaviour and stability in Syria (SEEDS)' (Brück)	International Security and Development gCmbH (ISDC)	Syria	Study completed Phase II <ul style="list-style-type: none"> <li>Presentation on the study in the Home Gardens for Resilience and Recovery Webinar Series</li> <li>Story Map of the SEEDS for Recovery project <a href="https://storymaps.arcgis.c">https://storymaps.arcgis.c</a></li> </ul>	Evaluation			

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
			<p><a href="https://cedilprogramme.org/stories/1392deed227f45a98e2093c922a03798">om/stories/1392deed227f45a98e2093c922a03798</a></p> <ul style="list-style-type: none"> <li>• CEDIL evidence brief: 'Syria: Do Complex Agricultural Interventions Strengthen Food Security' (Baliki et al. 2021)</li> <li>• <a href="https://cedilprogramme.org/events/seeds-for-recovery-the-impact-of-agricultural-interventions-in-syria/">Webinar: SEEDS for Recovery: The impact of agricultural interventions in Syria</a></li> <li>• Blog: 'How to make impact evaluations work in humanitarian and conflict settings' <a href="https://cedilprogramme.org/blog/how-to-make-impact-evaluations-in-humanitarian-and-conflict-settings-work/">https://cedilprogramme.org/blog/how-to-make-impact-evaluations-in-humanitarian-and-conflict-settings-work/</a></li> <li>• Final report: Impact evaluation of complex</li> </ul>				



CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
			agricultural interventions in Syria: An Overview. (Baliki et al., 2023).				
'A mixed-methods synthesis to develop a mid-range theory for interventions aiming to generate demand for contraception among adolescents' (Burchett)	London School of Hygiene and Tropical Medicine	Global	<p>Study completed Phase II</p> <ul style="list-style-type: none"> <li>• CEDIL research design paper: 'Upstream interventions aiming to encourage adolescents' use of contraception in low- and middle-income countries: A rationale and protocol for a mixed-methods synthesis to develop a mid-range theory'</li> <li>• Blog: <a href="#">"But HOW does it work?" Innovating family planning interventions</a></li> <li>• CEDIL working paper 5: <a href="#">'Structural interventions aiming to enable adolescent use of contraception in low- and middle-income countries'</a></li> </ul>	Evidence synthesis			

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
'Gender-Sensitive Risks and Options Assessment for Decision Making (ROAD) to Support WiF2' (Ringler)	International Food Policy Research Institute	Bangladesh Jordan Lebanon	<p>Study completed Phase II</p> <ul style="list-style-type: none"> <li>• CEDIL Design Paper 4: 'Gender-Sensitive Risks and Options Assessment for Decision Making (ROAD) to Support WiF2', design paper 4</li> <li>• 'The role of social identity in improving access to water, sanitation and hygiene (WASH) and health services: Evidence from Nepal', Research paper, 19 September 2021, <i>Development Policy Review</i>. <a href="https://doi.org/10.1111/dpr.12588">https://doi.org/10.1111/dpr.12588</a></li> <li>• '<a href="#">Reducing Vulnerability and Precarity of Low-Skilled Women in Short-Term Migration from the Global South: Key Policy Recommendations for the G-20</a>' – policy note on key risks along the migration pathway, on the role of women's empowerment</li> </ul>	Evaluation			

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
			<p>and on impact of WiF interventions, September 2021</p> <ul style="list-style-type: none"> <li>Final report not available in time for analysis</li> </ul>				
'Impact Evaluation of the SHARPE Project in Ethiopia' (de Brauw)	International Food Policy Research Institute	Ethiopia	<p>Study completed Phase II</p> <ul style="list-style-type: none"> <li>CEDIL Design Paper 7: 'Impact evaluation of Ethiopia's SHARPE program'</li> <li>Blog post: '<a href="#">Designing a community referral system to increase the financial access of refugee populations in Ethiopia</a>'</li> <li>Final report in peer review</li> </ul>	Evaluation			
'Gender and social outcomes of WASH interventions: synthesis of research evidence' (Macura)	The Stockholm Environment Institute	LMICs	<p>Study completed Phase II</p> <ul style="list-style-type: none"> <li>'<a href="#">Gender and social equality in WASH</a>' - blog post</li> </ul>	Evidence synthesis			

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
			<ul style="list-style-type: none"> <li>• <a href="#">‘Covid-19 and WASH’</a> – blog post</li> <li>• Final report being published</li> </ul>				
‘Involving men and boys in family planning: A systematic review of complex interventions to identify effective programme components and characteristics in low- and middle-income countries’ (Lohan)	Queen’s University Belfast	Global	<p>Study completed Phase II</p> <ul style="list-style-type: none"> <li>• <a href="#">‘Involving Men and Boys in Family Planning (Involve FP) in low- and middle-income countries’</a> – CEDIL webinar, 23 February, 2022</li> <li>• Protocol in Campbell Systematic Reviews, <a href="#">17(1), e1140</a>.</li> <li>• <a href="#">‘Involving men and boys in family planning: A systematic review of the effective components and characteristics of complex interventions’</a>, CEDIL research project paper 3</li> <li>• <i>This paper has also been published as a Campbell Systematic Review and is also available <a href="#">here</a></i></li> </ul>	Evidence synthesis			

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
'Evaluating portfolio interventions' (Masset)	Centre of Excellence for Development Impact and Learning	Not applicable		Conceptual paper (excluded)			
'Evaluation and measurement' (Attanasio)	Yale University	Not applicable		Conceptual paper (excluded)			
'Technology-based innovative solutions for improving perinatal care utilisation: a network meta-analysis' (Rahman)	Department of Global Health Policy, The University of Tokyo	Multi-country	<p>No longer supported by CEDIL due to UK aid cuts during COVID-19</p> <ul style="list-style-type: none"> <li>• CEDIL Design Paper 5: <a href="#">'Technology-Based Innovative Solutions for Improving Perinatal Care Utilization: a Network Meta-Analysis, Design paper 5'</a></li> </ul>	Evidence synthesis			
'Impact of Maternal Depression Treatment on Maternal Health, Parental Investment, and Child Development' (Maselko)	University of North Carolina at Chapel Hill, Gillings School of Global Public Health	Pakistan	<p>No longer supported by CEDIL due to UK aid cuts during COVID-19</p> <ul style="list-style-type: none"> <li>• CEDIL Design Paper 2: 'Impact of Maternal Depression Treatment on Maternal Health, Parental</li> </ul>	Evaluation			

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
			Investment, and Child Development’ <ul style="list-style-type: none"> <li>• Blog post: ‘Measuring patience among young children in rural Pakistan: Lessons learned from the marshmallow test’</li> <li>• Blog post: ‘COVID in Pakistan: its effects on field research work and future implications’</li> </ul>				
‘Evidence synthesis: the effectiveness of climate-related aid’ (Vanhuysse)	The Stockholm Environment Institute	Multi-country	No longer supported by CEDIL due to UK aid cuts during COVID-19	Evidence synthesis			
‘Evaluating governance reform using a case-control approach’ (Handa)	Carolina Population Center, University of North Carolina at Chapel Hill	Nepal	No longer supported by CEDIL due to UK aid cuts during COVID-19	Evaluation	x	x	x
<b>Programme of work 2: Enhancing evidence transferability</b>							
‘Making predictions of programme success more reliable’ (Cartwright)	Durham University	Not applicable	Study completed Phase I <ul style="list-style-type: none"> <li>• Cartwright, N., Charlton, L., Juden, M., Munslow, T.</li> </ul>	Conceptual paper (excluded)			



CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
			<p>and Williams, R.B. (2020) <a href="#">‘Making predictions of programme success more reliable’</a>. CEDIL Methods Working Paper. Oxford: CEDIL.</p> <ul style="list-style-type: none"> <li>• Cartwright, N. (2020) <a href="#">‘Using middle-level theory to improve programme and evaluation design’</a> CEDIL Methods Brief. Oxford: CEDIL.</li> </ul>				
‘Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (7-year follow-up)’ (Gertler)	Educate!/UC Berkeley	Uganda, Pakistan	<p>Study not yet completed</p> <ul style="list-style-type: none"> <li>• Design paper: ‘Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (8-year follow-up)’.</li> <li>• Working paper: <a href="#">‘Making Entrepreneurs: Returns to Training Youth in Hard Versus Soft Business Skills’</a> (presented at <a href="#">Online BREAD conference</a>)</li> </ul>	Evaluation	✓		

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
			<p><a href="#">on the economics of Africa</a> (7–9 July 2021) – hosted by the International Growth Centre and co-organised with the African Economic Research Consortium, the African School of Economics, and the Global Poverty Research Lab at Northwestern University)</p> <ul style="list-style-type: none"> <li>• Presentation: <a href="#">‘Making Entrepreneurs: Effect of Training Youth in Business Skills on Enterprise and Employment Creation’</a> (presentation in USC Economics Department, fall 2021 Development Seminar, 29 September)</li> </ul>				
‘Putting the theory of change to work: Process-Outcome Integration with Theory’ (Davey	London School of Hygiene and Tropical Medicine	Global	<p>Study completed Phase II</p> <ul style="list-style-type: none"> <li>• CEDIL design paper: <a href="#">‘POInT Research Design Paper, CEDIL Design Paper 3’</a></li> </ul>	Secondary data analysis			

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
			<ul style="list-style-type: none"> <li>Final report in peer review</li> </ul>				
'Using meta-analysis to explore the transferability of education mid-range theories to Cameroon, Chad, Nigeria and Niger' (Kay)	Education Endowment Foundation	Cameroon, Chad, Nigeria, Niger	<p>Study completed Phase II</p> <ul style="list-style-type: none"> <li>Blog: 'Building evidence infrastructure is a global good' (now removed from web)</li> <li>Presentation to FCDO on 14 February 2022: <a href="#">'Recontextualising the Teaching and Learning Toolkit to the Chad Basin For CEDIL'</a></li> <li>'Using meta-analysis to explore the transferability of education mid-range theories to Cameroon, Chad, Nigeria and Niger'. CEDIL Research project 2 (Higgins <i>et al.</i>, 2022)</li> </ul>	Evidence synthesis			✓
'Understanding factors that influence Teaching at the Right Level's effectiveness and generalizability: A Bayesian Evidence Synthesis' (Angrist)	Young 1ove Organisation	Global	<p>Study completed Phase II</p> <ul style="list-style-type: none"> <li>CEDIL synthesis paper 4: <a href="#">'The role of implementation in generalisability: A synthesis of evidence on</a></li> </ul>	Evidence synthesis		✓	

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
			<a href="#">targeted educational instruction and a new randomised trial'</a>				
'An empirically-driven theory of poverty reduction' (Handa)	Carolina Population Center, University of North Carolina at Chapel Hill	Malawi, Ghana, Zambia	Study completed Phase II <ul style="list-style-type: none"> <li>• Design paper 6: '<a href="#">An Empirically-driven Theory of Poverty Reduction'</a></li> <li>• Final report being published</li> </ul>	Secondary data analysis			
'Language Transitioning Research Synthesis' (Nakamura)	American Institutes for Research	Global	Study completed Phase II <ul style="list-style-type: none"> <li>• CEDIL Design Paper 12: 'Language of Instruction in Schools in LMICs: A Systematic Review'</li> <li>• Webinar: 'Language of Instruction Transitioning in Low-and- Middle Income Countries: An Evidence Synthesis'</li> <li>• Final report in peer review</li> </ul>	Evidence synthesis			
'Scaling Social Accountability for Health: Leveraging Public Policies and Programs' (Cant)	World Vision UK	Cambodia,	No longer supported by CEDIL due to UK aid cuts during COVID-19	Exploratory			

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
		Democratic Republic of Congo					
'Governance within forcibly displaced communities: evidence from Rohingya refugee camps' (Matin)	BRAC Institute of Governance and Development	Bangladesh	No longer supported by CEDIL due to UK aid cuts during COVID-19	Exploratory project			
'Syrian business development and regional trade in a humanitarian setting' (Baysan)	Harvard Business School HSBC Business School at Peking University	Syria	No longer supported by CEDIL due to UK aid cuts during COVID-19	Exploratory project			
'Catalysing responsive and inclusive governance: developing theory and methods for evaluating government training on citizen engagement' (Sheely)	Mercy Corps Europe	Myanmar and Jordan	No longer supported by CEDIL due to UK aid cuts during COVID-19	Exploratory project			
<b>Programme of work 3: Increasing evidence use</b>							
<ul style="list-style-type: none"> <li>'Engaging stakeholders with evidence and uncertainty' (Oliver)</li> </ul>	EPPI-Centre, University	Not applicable	Study completed Phase I <ul style="list-style-type: none"> <li>Oliver <i>et al.</i> (2021) 'Engaging Stakeholders'</li> </ul>	Conceptual paper (excluded)			

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
	College London		<p>with Evidence and Uncertainty: Developing a toolkit', CEDIL Methods Working Paper 4. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford.</p> <p><a href="https://doi.org/10.51744/CMWP4">https://doi.org/10.51744/CMWP4</a></p> <ul style="list-style-type: none"> <li>• CEDIL Methods Working Paper <a href="https://cedilprogramme.org/wp-content/uploads/2021/09/CEDIL-Methods-Working-Paper-4_Final.pdf">https://cedilprogramme.org/wp-content/uploads/2021/09/CEDIL-Methods-Working-Paper-4_Final.pdf</a></li> <li>• CEDIL Methods Brief <a href="https://cedilprogramme.org/wp-content/uploads/2021/09/CEDIL-Methods-Brief-5_Final.pdf">https://cedilprogramme.org/wp-content/uploads/2021/09/CEDIL-Methods-Brief-5_Final.pdf</a></li> <li>• <a href="#">View toolkit</a></li> </ul>				
<ul style="list-style-type: none"> <li>• Disability evidence portal (Shakespeare'</li> </ul>	London School of Hygiene and	Multi-country	<p>Study completed Phase I</p> <ul style="list-style-type: none"> <li>• <a href="#">Disability Evidence Portal</a></li> </ul>	Evidence portal			



CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
	Tropical Medicine						
<ul style="list-style-type: none"> <li>'Structural estimation of spatial spillover effects of cash transfers' (Gollin)</li> </ul>	University of Oxford	Kenya, Uganda, Lesotho, Tanzania	No longer supported by CEDIL due to UK aid cuts during COVID-19	Secondary data analysis			
<ul style="list-style-type: none"> <li>'How Biased are Observational Methods in Practice? Accumulating Evidence Using Randomised Controlled Trials with Imperfect Compliance' (Bryan)</li> </ul>	University of Warwick	Afghanistan, Kenya, Nepal	Study completed Phase II <ul style="list-style-type: none"> <li>Final report in peer review</li> </ul>	Evidence synthesis			
<ul style="list-style-type: none"> <li>'The art and science of using evidence: an evidence map, synthesis and evidence-informed guidelines for supporting the use of evidence in low- and middle-income countries' (Langer)</li> </ul>	Africa Centre for Evidence, University of Johannesburg	Low- and middle-income countries	No longer supported by CEDIL due to UK aid cuts during COVID-19	Evidence synthesis			
<ul style="list-style-type: none"> <li>'Predicting optimal policies for new contexts</li> </ul>	Development Research Institute, New	Multiple countries	No longer supported by CEDIL due to UK aid cuts during COVID-19	Evidence synthesis			

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

Project name and lead	Research team	Country	Outputs	Study type	Workshop invitation	Workshop presentation	Workshop survey
using existing studies' (Dehejia)	York University						
<ul style="list-style-type: none"> <li>'Enhancing evaluation use: rubric methodology in programme evaluation' (te Vogt)</li> </ul>	Mekong Economics Myanmar Ltd	Myanmar	<p>No longer supported by CEDIL due to UK aid cuts during COVID-19</p> <ul style="list-style-type: none"> <li><a href="#">CEDIL Design Paper 11: 'Enhancing evaluation use: An experimental assessment of rubric methodology on collective judgements in Myanmar'</a></li> </ul>	Exploratory project			

## Appendix 3

### Empirical claims made by CEDIL studies

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
Programme of work 1: Evaluating complex interventions			
<p><b>Evidence synthesis:</b> 'Involving men and boys in family planning: A systematic review of the effective components and characteristics of complex interventions in low- and middle-income countries' (Lohan)</p>	Effectiveness	<p><b>Report</b> (Aventin <i>et al.</i>, 2023): 'The most <b>effective interventions</b> are community-based educational programmes offered in schools, communities and homes or community facilities, and interventions involving multiple components, delivered by professionals, trained facilitators or peers to both males and females for over seven months. Brief programmes of less than three months are also effective.' This conclusion comes from standard methods for quantitative analysis: using <i>r</i> with robust variance estimation (RVE), meta-analysis and meta-regression.</p>	Explicit: findings from quantitative analysis, with methods described.
	Influential contextual factors	<p>Also important is promoting gender-equitable attitudes and social norms for women and girls among men and women at the individual, wider family, community, health service and societal level as part of family planning programming. Some studies also emphasised structural factors, such as the importance of widening women's access to education and labour markets. These claims were justified by qualitative analysis involving framework synthesis framed by a logic model of intervention</p>	Explicit: findings from implementation studies, with methods of framework synthesis described, including a priori and emergent themes.

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
	<p>Values and norms</p>	<p>components based on stakeholder consultation and prior research.</p> <p>‘The success of family planning programmes that involve men and boys is most often measured by contraceptive use to the relative neglect of other outcomes, such as met need for family planning, equitable family planning decision-making, or gender equality.’ Justified by comparing the description of studies to the logic model.</p> <p>‘Our qualitative analysis also highlights the under-used strategy of addressing gender equality attitudes and norms, from the individual to the structural level.’</p> <p>The causal chain was unclear: ‘across the range of proximal and distal outcome measures (including contraceptive use, desired family size, pregnancy, pregnancy timing, gender equitable attitudes, communication about FP, equitable decisionmaking about FP, attitudes about FP, knowledge about contraceptives, and FP service use) there were few clear or consistent findings.’</p>	<p>Explicit: description of studies compared with logic model.</p> <p>Explicit: description of studies compared with logic model.</p> <p>Explicit: description of studies compared with logic model.</p>

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
<p><b>Evidence synthesis:</b> ‘A mixed-method synthesis to develop a mid-range theory for interventions aiming to generate demand for contraception among adolescents’ (Burchett)</p>	<p>Identifying key elements for interventions and implementation</p> <p>Local and age group relevance due to stakeholder engagement</p> <p>Factors explaining variance</p> <p>Mid-range theory</p>	<p><b>Design paper</b> (Burchett, 2022): This study will ‘explore... what intervention characteristics may facilitate or hinder their effectiveness’ (p1)... to develop ‘a framework that sets out the key elements (e.g. characteristics of the intervention content, context or implementation) that should be incorporated into interventions and their implementation’ (p. 2).</p> <p>Backed up by local stakeholder engagement in the form of advisory groups – one for policymakers, non-governmental organisation staff and academics, and one with adolescents in Mozambique who will be consulted to refine the theory proposed based on findings (Burchett, 2022 p. 8).</p> <p><b>Final report</b> (Burchett <i>et al.</i>, 2022): Claims about which <b>intervention factors</b> were associated with effectiveness are not made due to limitations in the evidence.</p> <p>A <b>mid-range theory</b> for contraceptive use interventions, requiring six objectives to be met was developed using ICA. Three objectives were strongly influenced by upstream factors that are likely to be best addressed by structural interventions: desire to</p>	<p>Explicit; more detailed text (p. 4) explains the potential of QCA and ICA. However, quality appraisal methods for possibly excluding low quality studies were not defined in advance.</p> <p>Implicit: assumption that the participation of the advisory groups contributes to the validity of the final results.</p> <p>Explicit: ‘We conducted an evidence synthesis with a comprehensive systematic search and</p>





CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
	<p>Equity: knowledge gaps</p>	<p>empowerment, and women’s participation and leadership in WASH services.</p> <p>Only about half of data were disaggregated across age, gender, and/or social category; most often, women and girls; rarely on a specific caste or class, the elderly, ethnicity, religion, or people with disabilities; and never on gender and sexual minorities.</p> <p>Most interventions were designed without a specific gender equity and social inclusion (GESI) component, and few evaluated outcomes related to gender-based violence and other forms of violence against marginalised groups in the WASH sector.</p> <p>Most studies focused on households and schools; research is needed in other settings, such as healthcare facilities and workplaces. Most studies were conducted in certain geographical regions, such as India and Kenya; however, the importance of social and cultural drivers for understanding GESI outcomes requires more clarity on these dynamics in other geographical settings.</p>	

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
	Context: knowledge gaps		
<p><b>Evidence synthesis:</b> ‘Technology-Based Innovative Solutions for Improving Perinatal Care Utilization: A Network Meta- Analysis’ (Rahman)</p>	Knowledge gap	<p><b>Design paper</b> (Rahman <i>et al.</i>, 2022): ‘Critically, [no studies] assess the complex interplay and impact of technology-based interventions over the dynamic antenatal-delivery-postnatal pathway’ (p. 6).</p> <p>This systematic review and network meta-analysis aimed to fill this gap in knowledge. It adopted well-established systematic review methods and tools.</p>	<p>Justification of the gap was supported by citations of available studies, although the absence of a systematic review was not stated.</p> <p>Justification that the study will fill the gap was implicit in the choice of methods and tools.</p>
<p><b>Evaluation:</b> ‘Gender-Sensitive Risks and Options Assessment for Decision Making (ROAD) to Support WiF2’ (Ringler)</p>	Effects of intervention	<p><b>Design paper</b> (Abdulrahim, 2022): Explicit claims about the effects of interventions were avoided by adopting the objective ‘to explore the impact of technology-based interventions in improving ANC visits, PNC visits, and delivery care services in LMICs using Bayesian network meta-analysis’ (p. 5).</p> <p><b>Final report:</b> Not available in time for analysis.</p>	Justification implied by use of established methods for systematic review and network meta-analysis.

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
<p><b>Evaluation:</b> ‘Seeds for recovery: The long-term impacts of a complex agricultural intervention on welfare, behaviour and stability in Syria (SEEDS)’ (Brück )</p>	<p>Effects at community level</p> <p>Effects and equity at household level</p> <p>Effects of context</p>	<p><b>Final report</b> (Baliki et al. 2023): Irrigation rehabilitation interventions improved community agricultural productivity, but only in the spring season.</p> <p>Female-headed households benefitted remarkably more in terms of food security in the medium term compared to male-headed families.</p> <p>Moreover, households residing in areas that were moderately affected by violent conflict showed stronger food security improvements compared to households from peaceful or conflict-intense settings.</p> <p>Justified by innovative methods for collecting data via satellite photography, and alternative methods for constructing counterfactuals.</p>	<p>Explicit.</p>
<p><b>Evaluation:</b> ‘Impact Evaluation of the SHARPE Project in Ethiopia’ (de Brauw)</p>	<p>Factors supporting or constraining interventions</p>	<p><b>Design paper</b> (de Brauw 2022): Claim that the study would identify these factors is supported by examples of constraints already identified.</p> <p>‘Qualitative work (complemented by quantitative work) would aim to understand what actual support factors and derailers are relevant in the study contexts, as well as whether intermediate steps are</p>	<p>Justification for identifying factors was explicit.</p>

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
	Effects of intervention	<p>occurring between tested strategies and targeted objectives’ (p. 13).</p> <p><b>Final report</b> (de Brauw et al. 2023): Empirical claims were limited. As things turned out, local staff losses following political conflict, and COVID travel restrictions, affected the research team’s ability to build a local understanding of the context, to work on relationships needed to manage the project, and to solve problems from a local perspective</p>	The role of randomisation in justifying evidence of effects was implicit (see methodological claims below about rapid RCTs).
<p><b>Evaluation:</b> ‘Impact of Maternal Depression Treatment on Maternal Health, Parental Investment, and Child Development’ (Maselko)</p>	<p>Burden of illness</p> <p>Effect of transitions</p> <p>Improved interventions</p> <p>Pathways from adversity to child development</p> <p>Effects of intervention</p>	<p><b>Design paper</b> (Maselko <i>et al.</i>, 2022): Claims to generate empirical evidence rested on claims relating to methodological innovation (see Appendix 4).</p> <p>‘We will estimate intention to treat (ITT) effects for multiple outcomes, following established best-practice statistical methods in impact evaluation (e.g., randomization inference accounting for clustering, adjustment for multiple comparisons,</p>	<p>The strength of links between empirical claims and methodological claims varied.</p> <p>Justification implied by choice and details of established methods</p>

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
	Local relevance as data collection carried out by trained suitable researchers	<p>attrition bounds) (Duflo, Glennerster, and Kremer 2007). We will investigate if attrition is differential by treatment status... to assess robustness of the estimates to any differences' (p. 15).</p> <p>"Members of the study team have been working in this area for over 20 years, building strong relationships with the mothers and families", "Data are collected by female research assistants with robust training who thoughtfully attend to the cultural context of the area" (p. 4)</p> <p>No final report because no longer supported by CEDIL due to UK aid cuts during COVID-19.</p>	Implicit assumption that this particular method of data collection yields 'better' findings.
<b>Programme of work 2: Enhancing evidence transferability</b>			
<b>Evidence synthesis:</b> generalizability: a Bayesian evidence synthesis' (Angrist)	Effectiveness Fidelity	<p><b>Final report</b> (Angrist and Meager 2022): Standard statistical meta-analysis methods justified claims about effectiveness and fidelity:</p> <ul style="list-style-type: none"> <li>Reported effectiveness of targeted instruction varied from 0.07 to 0.78 standard deviations (SDs) across contexts. Two factors explained most of the heterogeneity in reported effects: the degree of implementation (intention-to-treat or treatment-on-the-treated effects) and the</li> </ul>	Explicit.

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
		<p>instruction delivery model (teachers or volunteers).</p> <p>A RCT justified the higher effects with high fidelity:</p> <ul style="list-style-type: none"> <li>Targeted instruction can deliver 0.39 SD improvements in learning on average when taken up, and 0.80 SD gains when implemented with high fidelity, explaining the upper range of effects in the literature. Increasing programme fidelity led to additional 0.22 SD gains relative to standard implementation.</li> </ul>	<p>Explicit.</p>
<p><b>Evidence synthesis:</b> 'Using meta-analysis to explore the transferability of education mid-range theories to Cameroon, Chad, Nigeria and Niger' (Kay)</p>	<p>Effectiveness</p>	<p><b>Final report</b> (Higgins <i>et al.</i>, 2021): Systematic reviews found:</p> <ul style="list-style-type: none"> <li>a small positive effect on attainment from cash transfers, but assuming that providing cash without monitoring or enforcing attendance would increase pupil attendance and attainment at school was not well-supported by existing evidence;</li> <li>too few studies to meta-analyse for an attainment outcome of menstrual hygiene interventions; and</li> <li>no rigorous studies had evaluated the effects of corporal punishment.</li> </ul>	<p>Explicit: findings based on Standard Education Endowment Foundation systematic review methods.</p>

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
	<p>Relevance of evidence</p>	<p><i>Judged by stakeholders:</i> Policymakers and practitioners within the Chad Basin judged the relevance of the pedagogical approaches included in the Toolkit to be largely appropriate and applicable in their context. Five approaches (one-to-one tuition, school uniforms, setting and streaming, teaching assistant interventions, and learning styles) were considered either too expensive or otherwise infeasible to implement in the context. The other three topics (menstrual hygiene interventions, cash transfers, and corporal punishment) were received enthusiastically by policymakers and school leaders.</p> <p><i>Judged by geography:</i> In contrast to studies in high-income countries, studies LMICs are much more likely to focus on policy levers (e.g. performance pay) than pedagogical strategies.</p> <p>Despite the relevance of pedagogical approaches to the Chad Basin, relatively few research studies have examined pedagogical approaches, such as feedback or metacognition to improving learning outcomes.</p> <p>The contrast between studies in metacognition, for example: there have been 241 studies that have taken place in high-income countries and 12 that</p>	<p>Explicit: expert group described and named; questions put to them listed.</p> <p>Comparing studies in Education Endowment Foundation toolkit with studies in 3ie repository.</p>

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
	<p>Research priorities</p> <p>Values and norms</p>	<p>have taken place in LMICs. Many promising pedagogical strategies in the Toolkit have never been evaluated in a LMIC context.</p> <p>In both consultations with teachers and policymakers, there was a consistent desire to include information on cash transfers and menstrual hygiene interventions. These topics were frequently cited as interventions considered by policymakers to improve educational outcomes.</p> <p>The other topic area that was frequently discussed was the use of corporal punishment as a behaviour management approach. Stakeholders did not see the approach as a desirable intervention and the United Nations Convention on the Rights of the Child and African Charter on the Rights and Welfare of the Child have called for an end to corporal punishment. Nevertheless, it is still widely practised. Policymakers requested better evidence so that they could advocate for other interventions.</p>	
<p><b>Evidence synthesis:</b> 'Language transitioning research synthesis' (Nakamura)</p>	<p>Effects of interventions</p>	<p><b>Final report</b> (Nakamura, Molotsky <i>et al.</i>, 2022): Mother tongue-based language of instruction interventions 'may improve students' letter knowledge, word reading, sentence reading, and reading comprehension in the students' MT, improve</p>	<p>Implicit.</p>



CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
		<p>students' word and sentence reading, and reading comprehension in the national language, and improve students' oral language proficiency, word, and sentence reading, reading comprehension, and writing in the later acquired language'. Justified by meta-analyses and quantitative narrative synthesis.</p> <p>'High-quality teaching and learning materials in the MT coupled with improved curriculum and bilingual materials throughout the classroom are necessary for a successful MT-based LOI program. These programs overwhelmingly received positive reception and wide support by students and teachers alike as students and parents perceive that these programs improve teaching quality, increase student motivation in the classroom, and respondents report improvements in bilingual reading skills.' Justified by 'best-fit' framework synthesis'.</p>	
<p><b>Evaluation:</b> 'Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (7-year follow- up)' (Gertler)</p>	<p>Knowledge gap and importance of study</p>	<p><b>Design paper</b> (Chioda and Gertler, 2022): Findings from employer surveys (2010) and an ongoing systematic review about business training programmes justified the focus on (1) which (combination of) skills are important for leadership and entrepreneurship and how to teach them; (2) the sustainability of these interventions' impacts; (3) documenting any spillovers beyond the usual economic outcomes, such as risky behavior and IPV;</p>	

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
	Assertion that the findings will be useful to policymakers	<p>(4) identifying the mechanisms at play underlying effects, as well as the sub-groups that are most likely to benefit from these types of programmes.</p> <p>Use of regression trees and machine learning methods will produce findings that ‘will offer direct guidance to policy makers as to program design and policy targeting while abstracting from program-specific features to identify which set of skills and/or program components are effective levers such that, if activated by other programs, they could move outcomes along the causal chain’ (p. 23).</p> <p><b>Final report:</b> Study not yet completed.</p>	Implicit, need to know how machine learning methods work to know the benefits.
<b>Evaluation:</b> ‘An Empirically-driven Theory of Poverty Reduction’ (Handa )	Assessing impact	<b>The design paper</b> (Handa 2022) named four national unconditional cash transfer programmes and the study designs used to evaluate their impact: RCT (in Malawi and Zambia); discontinuity design using the proxy means test cut-off (in Ghana); and matched Wards (administrative units below the district), followed by the application of household targeting by the programme so all households in the comparison Wards are future eligible households (in Zimbabwe).	Justification was implicit, with no comment on the choice of different designs or the formative research that led to their development.

CEDIL project and lead	Q1: What are claims about?	Q2: What are the empirical claims, and how are they justified?	Q3: Is the justification explicit or implicit?
	<p>Effects Equity/ moderators</p>	<p>All evaluations included one baseline and multiple follow-ups (except for Ghana, with just one follow-up).</p> <p><b>Final report</b> (Handa <i>et al.</i>, 2022): Made a series of empirical claims about programme success, composition of more/ less successful households, consumption and assets of more/ less successful households. Empirical claims were compared with the poverty trap theory, which predicts that households can escape poverty with a big push, in order to overcome technological indivisibilities and other market frictions. This comparison suggests that only a particular group of the ultra-poor can escape poverty through unconditional cash transfer alone: those households at the younger side of the lifecycle, and who are provided with transfers over a long time period.</p>	<p>Individual claims were explicit and supported by empirical analysis. These were combined, and compared with existing theory, to refine that theory.</p>

## Appendix 4

### Methodological claims made by CEDIL studies

CEDIL project and lead	Q1: What are claims about?	Q2: What are the methodological claims and how are they justified?	Q3: Is the justification explicit or implicit?
<b>Programme of work 1: Evaluating complex interventions</b>			
<b>Evaluation:</b> 'Gender-Sensitive Risks and Options Assessment for Decision Making (ROAD) to Support WiF2' (Ringler)	Mixed-methods approach will combine benefits of both and therefore produce reliable findings	<p><b>Design paper</b> (Abdulramin 2022): 'The ROAD process follows a 'mixed-methods' approach for evaluation of risks across all levels of governance and decision-making. It encompasses complementary quantitative and qualitative phases, each integrating multi-disciplinary strengths' (p. 6).</p> <p><b>Final report:</b> Not available in time for analysis.</p>	Implicit justification: mentioned 'strengths' but without going into specifics.
'Seeds for recovery: The long-term impacts of a complex agricultural intervention on welfare, behaviour and stability in Syria (SEEDS)' (Brück)	<p>In areas of conflict:</p> <p>Innovative data collection</p> <p>Innovative data analysis</p>	<p><b>Final report</b> (Baliki et al., 2023):</p> <p>(a) Multidimensional and longitudinal household panel survey dataset from treatment and control groups, to analyse evolution of effects over time.</p> <p>(b) High-resolution satellite data to examine the impact of the intervention at the village levels.</p> <p>Constructing counterfactuals using adaptive research designs and applying flexible matching and balancing to improve sample comparability allow rigorous impact evaluation.</p> <p>Supervised machine learning and deep learning overcome challenges in traditional data and impact evaluation designs, such as small sample sizes and assignment imbalances. Machine learning</p>	Explicit demonstration of methods.

CEDIL project and lead	Q1: What are claims about?	Q2: What are the methodological claims and how are they justified?	Q3: Is the justification explicit or implicit?
		<p>techniques, such as the causal random forest, allow for exploring the data in depth and conducting impact heterogeneity analysis without losing statistical power.</p> <p>Using and combining various types of data, such as remote-sensed data, conflict event data, and household survey data, provides clear benefits.</p> <p>All justified in this demonstration project, except the last.</p>	
<p><b>Evidence synthesis:</b> ‘A mixed-method synthesis to develop a mid-range theory for interventions aiming to generate demand for contraception among adolescents’ (Burchett, 2022)</p>	<p>Valid choice of methods</p>	<p><b>Design paper</b> (Burchett 2002): They claimed they would use QCA ‘to consolidate existing theory and empirical research and... set out the principles and mechanisms through which a category of interventions can achieve outcomes for a particular issues. These principles and mechanisms can then be used to develop specific effective interventions that are appropriate for particular populations and settings’ (p. 4).</p> <p>This was justified ‘By using ICA, [they] will be able to capture both explicit, name theories used to develop interventions, but also theories implicit in the development of interventions, and/or in explanations of their results. [They] will be able to</p>	<p>Explicitly explained the uses of QCA and ICA, as well as the added value of combing them (section ‘Exploration of the value of ICA and QCA as tools for developing mid-range theory’, p. 6).</p>

CEDIL project and lead	Q1: What are claims about?	Q2: What are the methodological claims and how are they justified?	Q3: Is the justification explicit or implicit?
		<p>explore the value of these novel evidence sources’ (p 5).</p> <p>This was justified because ‘QCA may be particularly useful for generating mid-range theory as it allows an exploration of some of the complex causal relationships that take place in the natural world but that can otherwise be challenging to identify and measure’ (p. 6).</p> <p>They claimed they would make ‘use of stakeholder engagement to shape the development of the mid-range theory’ (p. 7). This was only partially justified because ‘there is much left to be understood around how the involvement of stakeholders can clarify the concepts represented, whether additional adverse impacts are accounted for with the input of stakeholders, how the involvement should be managed, and what happens when there are divergent views between the stakeholders and the evidence’ (p. 7) and they offer no solutions in the methods.</p>	<p>The claim about stakeholder engagement was implicit and was supported by citations: ‘Stakeholder engagement as critical in order to develop theories that are salient to decision-makers and intervention recipients (Kneale, Thomas <i>et al.</i> 2015)’ (p. 7).</p>
<p><b>Evidence synthesis:</b> ‘Gender and social outcomes of WASH interventions: synthesis of research evidence’ (Macura)</p>	<p>Conceptualising and measuring outcomes</p>	<p><b>Design paper</b> (Macura et al., 2022): ‘This review will help to conceptualize gender and social equality outcomes in the WASH work and support the development of tools for accurate measurement and evaluation of GSE outcomes. As a result, the review findings will equip policy-makers and practitioners</p>	<p>Statement of intent, rather than justifying a claim.</p>

CEDIL project and lead	Q1: What are claims about?	Q2: What are the methodological claims and how are they justified?	Q3: Is the justification explicit or implicit?
		with evidence for effective mainstreaming of gender and social equality in WASH interventions.'	
<b>Evidence synthesis:</b> 'Technology-Based Innovative Solutions for Improving Perinatal Care Utilization: A Network Meta-Analysis' (Rahman)		<b>Design paper</b> (Rahman et al. 2022): Implicit assumption that a systematic review following PRISMA guidelines, employing Cochrane Collaboration's tool and ROBIS-I to assess study quality, will answer the question.  No final report because no longer supported by CEDIL due to UK aid cuts during COVID-19.	Methodological choices were not justified; justification was implicit and relied on readers understanding the methods and tools
<b>Evidence synthesis:</b> 'Using big data for evaluating development outcomes: a systematic map' (Rathinam 2020)	Valid choice of methods	<b>Design paper</b> (Rathinam et al., 2020): Argument and existing studies supported the claims that big data can: <ul style="list-style-type: none"> <li>• measure development outcomes that are difficult to measure using household surveys or administrative data, such as economic output, wealth, population movement, or disease spread within a given local area;</li> <li>• identify comparison groups (because big data are generally available before and after programmes, and for programme areas and comparison sites);</li> <li>• measure long-term programme impacts and sustainability (because some sources of big data are available for a reasonably long period);</li> </ul>	Explicit justification: existing measurement studies served as a proof of concept for big data contributing to the evidence base in development sectors where evaluations are often infeasible due to data issues.  Explicit caveats: Very few studies report on big data ethical issues: informed consent, data privacy, data security

CEDIL project and lead	Q1: What are claims about?	Q2: What are the methodological claims and how are they justified?	Q3: Is the justification explicit or implicit?
		<ul style="list-style-type: none"> <li>• provide data on pre-programme trends and control variables for better statistical precision;</li> <li>• provide data to evaluate the different impacts on different sub-groups;</li> <li>• aid in conducting robustness analyses based on multiple comparison groups and placebo tests; and</li> <li>• evaluate the impact retrospectively and at the level of individual projects, as well as at programme/portfolio level.</li> </ul>	<p>and unintended exclusion.</p> <p>Several analytical, ethical and logistical challenges may hinder the use of big data in evaluations. The report called for standards to be set for the reporting of data quality issues, data representativeness and data transparency.</p>
<p><b>Evaluation:</b> ‘Impact Evaluation of Ethiopia’s SHARPE programme’ (de Brauw)</p>	<p>Appropriate combination of methods</p>	<p><b>Design paper</b> (de Brauw, 2022): Adding rapid RCTs to long-term projects to inform subsequent intervention adaptations was justified by carefully constructing an initial theory of change and applying primarily qualitative methods to identify intervention constraints which might be opportunities for intervention improvements.</p> <p><b>Final report</b> (de Brauw et al., 2023): Adding rapid RCTs turned out to be challenging: one was too complex both to explain to partners and to set up. Time constraints limited the collection of qualitative data to explain the disappointing findings.</p>	<p>The justification was explicit.</p>



CEDIL project and lead	Q1: What are claims about?	Q2: What are the methodological claims and how are they justified?	Q3: Is the justification explicit or implicit?
<p><b>Evaluation:</b> 'Impact of Maternal Depression Treatment on Maternal Health, Parental Investment, and Child Development' (Maselko)</p>	<p>Innovative combination of impact measures</p> <p>Innovation of biomarkers in RCT</p> <p>Unique study design: RCT nested in cohort study</p>	<p><b>Design paper</b> (Maselko <i>et al.</i>, 2022): Including 'novel biological markers to enable early detection and mechanistic identification of treatment effects, alongside an unusually rich set of measures of child development, maternal function and parental investments, using state of the art psychometric tools' (p. 7).</p> <p>'It is rare to have data on chronic HPA-axis dysregulation in large longitudinal studies with measurements at multiple time points, such as what we proposed in this study. It is unique to have such longitudinal biomarkers embedded into a randomized control trial' (p. 7).</p> <p>'gathering of longitudinal data in a birth cohort study within which prenatally depressed mothers were randomized into treatment for maternal depression. We also recruit women who were pregnant but not depressed at baseline as their outcomes provide an upper bound on treatment effects under plausible assumptions' (p. 7).</p>	

CEDIL project and lead	Q1: What are claims about?	Q2: What are the methodological claims and how are they justified?	Q3: Is the justification explicit or implicit?
	<p>Innovative analytic approach</p> <p>Innovative provision of feedback on child performance</p> <p>Machine learning methods as best choice of analysis tool-&gt; will yield reliable findings</p>	<p>'By using machine learning to examine heterogeneity of treatment effects, we will identify key variables that are relevant to intervention targeting' (p. 7).</p> <p>'We incorporate an experiment testing parental responses to the delivery of personalized feedback to parents about their child's performance during developmental testing' (p. 7).</p> <p>'Given that the current sample size might be too small to detect an interaction effect between the intervention and all of the baseline maternal characteristics that we can consider. Therefore, we decided to tie our hands by using predefined machine learning algorithms that automatically avoid p-hacking and control for multiple hypothesis testing' (p. 15).</p>	<p>Methodological choices were not justified; justification was implicit and relied on the understanding of the methods (e.g. knowing what p-hacking is).</p>
<b>Programme of work 2: Enhancing evidence transferability</b>			
<p><b>Evidence synthesis:</b> 'Understanding factors that influence Teaching at the Right Level's effectiveness and generalizability: A Bayesian Evidence Synthesis' (Angrist)</p>	<p>Reporting implementation</p>	<p><b>Final report</b> (Angrist and Meager, 2022): 'The results of our analysis demonstrate the importance of quantifying programme implementation with as much care as we typically apply when quantifying programme effects. We find that implementation levels and delivery modes explain most of the</p>	<p>Explicit.</p>

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

CEDIL project and lead	Q1: What are claims about?	Q2: What are the methodological claims and how are they justified?	Q3: Is the justification explicit or implicit?
		variation in effects of targeted instruction across settings and that this leads to actionable insights that can improve the effectiveness of a programme that is being scaled up.'	
<b>Evidence synthesis:</b> 'Using meta-analysis to explore the transferability of education mid-range theories to Cameroon, Chad, Nigeria and Niger' (Kay)	Judging transferability	<b>Final report</b> (Higgins <i>et al.</i> , 2021): There is value in comparing across the studies from high-income countries and LMICs.  Discussions with stakeholders and the cultivation of a shared database have been incredibly valuable for assessing transferability. In areas of limited evidence, promising strategies from high-income countries might represent best bets for careful recontextualisation and evaluation.	Explicit argument.
<b>Evidence synthesis:</b> 'Language of instruction in schools in LMICs: a Systematic Review' (Nakamura, )	Valid choice of methods	<b>Design paper</b> (Nakamura, Leyew <i>et al.</i> , 2022): Claimed the research questions about impact can be answered by studies meeting eligibility and quality criteria were implicit in the description of methods.	Implicit.
<b>Evaluation:</b> 'Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (8-year follow- up)' (Gertler)	Validity of measures.  Combination of methods will yield reliable findings	<b>Design paper</b> (Chioda and Gertler, 2022): Cited papers that validate instruments in context of interest (p. 12).  'Our work will combine the machine learning (i.e., generalized random forest) and causal mediation analysis literatures to go beyond the "effect of a cause" (i.e., the treatment effect) and investigate the	Explicit, by citing earlier papers.  Implicit rather than explicit, because relied on the reader's understanding of the strengths and limitations of the





CEDIL project and lead	Q1: What are claims about?	Q2: What are the methodological claims and how are they justified?	Q3: Is the justification explicit or implicit?
	<p>Mentioned the potential limitations that might arise from using pre-existing data from evaluations, but negated by mentioning the high quality of these case study evaluations</p>	<p>teams. Three of the four studies are cluster randomised trials, with little risk of selection issues or spillovers. The quality of the qualitative work is also likely to be high, with well-resourced process evaluations being completed by appropriate teams’ (p. 27). ‘Data quality should be high since the evaluations are all being conducted by world-class multidisciplinary teams’ (p. 30).</p> <p><b>Final report</b> (Juden <i>et al.</i>, 2023): This study developed a method for translating a theory of change into a directed acyclic graph (or DAG) and applying Bayesian causal modelling to help evaluators make better inferences about the theory of change and the transportability of the results to other contexts. This was justified by applying the methods to two development interventions, and reflecting on the successes, challenges and limitations.</p>	<p>The choice of methods was very implicit: methodology was presented in a descriptive manner, without any explanation of the complex series of decisions taken which can be understood by a non-expert. However, explicitly assumed confidence in future findings by referring to the high level of quality of input data.</p> <p>Explicit.</p>
<p><b>Evaluation:</b> ‘An Empirically-driven Theory of Poverty Reduction’ (Handa )</p>	<p>Application of innovative methods</p>	<p><b>Design paper</b> (Handa, 2022): This study claimed to innovate by adopting ‘recent developments at the intersection of machine learning (ML) and causal inference to build a “middle-range” theory around poverty traps... indicating which households are more likely to benefit from a cash transfer versus other types of interventions, and what the</p>	<p>This claim was justified explicitly by a detailed argument.</p>

CEDIL project and lead	Q1: What are claims about?	Q2: What are the methodological claims and how are they justified?	Q3: Is the justification explicit or implicit?
		<p>pathways are to sustained increases in consumption’.</p> <p><b>Final report</b> (Handa <i>et al.</i>, 2022): ‘We have used causal forests, a relatively new machine learning algorithms, to identify heterogeneous treatment effects in four national UCTs in Africa. The advantage of this approach is that it allows the data to identify organically the high-flyers, households who are most sensitive to the intervention.’</p>	<p>Technical details were provided in the appendix.</p>
<b>Programme of work 3: Enhancing use of evidence</b>			
<p><b>Evidence synthesis:</b> ‘How Biased are Observational Methods in Practice? Accumulating Evidence Using Randomised Controlled Trials with Imperfect Compliance’ (LSE) (Bernard)</p>	<p>Importance of randomisation</p>	<p><b>Final report</b> (Bernard <i>et al.</i>, 2023): ‘Observational methods have mean zero bias, but the bias distribution has high variance such that any given observational study is likely to have high bias.’</p> <p>This claim was justified by reanalysing RCTs with incomplete compliance to compare them: intention-to-treat with treatment was accepted.</p>	<p>The claim was justified explicitly, but this involved an over-interpretation of the data because it ignored observational methods that construct control groups retrospectively.</p>
<p><b>Evaluation:</b> ‘Rubric methodology in programme evaluation’ (te Vogt)</p>	<p>Rubric methodology</p>	<p><b>Design paper</b> (Peterson <i>et al.</i>, 2022): ‘Our findings will highlight how rubric methodology transfers across evaluation contexts and provide insight on the conditions in which rubric methodology may be a useful approach for evaluation practitioners’.</p>	<p>Explicit, because it was being tested in Myanmar, although this is only one context.</p>

## Appendix 5

### Communicating claims made by CEDIL studies

CEDIL project and lead	What are claims about?	How is the confidence in the claim communicated for different audiences?	Can confidence in the claim communicated be traced back to the original assessments?
<b>Programme of work 1: Evaluating complex interventions</b>			
<p><b>Evidence synthesis:</b> 'Involving men and boys in family planning: A systematic review of the effective components and characteristics of complex interventions in low- and middle-income countries' (Aventin)</p>	<p>Logic model</p>	<p><b>Main report:</b> The logic model was revised: all information that was not evidenced (i.e. not significant or not included) in the included evaluation studies and connected papers was changed from black to grey font, to highlight areas for future research to consider.</p>	
<p><b>Evaluation:</b> 'Gender-Sensitive Risks and Options Assessment for Decision Making (ROAD) to Support WiF2' (Ringler)</p>	<p>Importance of topic</p>	<p><b>Policy note:</b> Both the design paper and the policy note used similar language to communicate the issue of female migration and forced labour, and provided a similar level of depth of explanations.  <a href="https://www.t20italy.org/2021/09/22/reducing-vulnerability-and-precarity-of-low-skilled-women-in-short-term-migration-from-the-global-south-key-policy-recommendations-for-the-g-20/">(https://www.t20italy.org/2021/09/22/reducing-vulnerability-and-precarity-of-low-skilled-women-in-short-term-migration-from-the-global-south-key-policy-recommendations-for-the-g-20/)</a></p>	<p><b>Policy note:</b> While not directly referencing the study, the policy recommendation to G20 also made reference to Sustainable Development Goal 8.7, like the design paper, to justify the study's relevance and its importance in a global context.</p> <p><b>Research paper:</b> No link to the design paper study.</p>



CEDIL project and lead	What are claims about?	How is the confidence in the claim communicated for different audiences?	Can confidence in the claim communicated be traced back to the original assessments?
<p><b>Evaluation:</b> ‘Seeds for recovery: The long-term impacts of a complex agricultural intervention on welfare, behaviour and stability in Syria (SEEDS)’ (Brück)</p>	<p>Prevalence</p>	<p>Blog <b>post:</b> Some claims in the blog were justified with images, graphs or links to academic papers. For instance, <b>this statement</b> about empirical findings – ‘Since the onset of the war, agricultural production in Syria has dropped severely, with millions in rural areas requiring emergency assistance and support’ – was supported by direct access to data in the form of Google Earth historical imagery which shows that more recent photographs are less green. Confidence in the data relied on readers’ interpretation of the images. Similarly, this methodological claim was partially justified by interactive images: ‘We use the Enhanced Vegetation Index (<b>EVI</b>), along with other vegetation indices, to measure vegetative vigor. EVI corrects for atmospheric influences and soil background signals in areas with dense vegetation.’</p> <p>An evidence brief provided an easy to read summary of the key findings and lessons learnt, with a histogram and a forest plot. A very brief methodological note at the end of the evidence brief mentioned the challenges of a panel survey and the</p>	<p>The blog provided links to underpinning publications, although these were not necessarily funded by CEDIL. For instance, the statement – ‘A significant reduction in irrigated area was found by Jaafar et. al (2015), who developed a model using remote sensing and GIS to assess the impact of the Syrian conflict on agriculture’ – was supported by a direct link to the <a href="#">publication</a>.</p>

CEDIL project and lead	What are claims about?	How is the confidence in the claim communicated for different audiences?	Can confidence in the claim communicated be traced back to the original assessments?
		<p>limitations imposed by late changes in beneficiary villages and households due to post-baseline context analysis recommendation conducted by the United Nations Food and Agriculture Organization.</p> <p>A blog described in plain language how to make impact evaluations work in humanitarian and conflict settings.</p>	<p>The evidence brief and the blog are hosted on the CEDIL website and can be accessed from the same home page as the full report.</p>
<p><b>Evidence synthesis:</b> 'Upstream interventions aiming to encourage adolescents' use of contraception in low- and middle-income countries' (Burchett, 2022)</p>	<p>Knowledge gap</p> <p>Appropriate methods</p>	<p><b>Blog post:</b> 'We know that reducing adolescent child-bearing is a global priority and that contraception is one of the ways of doing this. There is a lot of evidence on the effectiveness of interventions to increase the use of contraception.' Prior assumptions were communicated as an assertion; an expert opinion talking to a general non-expert audience (<a href="https://www.lshtm.ac.uk/research/centres/march-centre/news/217376/how-does-it-work-innovating-family-planning-interventions">https://www.lshtm.ac.uk/research/centres/march-centre/news/217376/how-does-it-work-innovating-family-planning-interventions</a>)</p>	<p><b>Blog post:</b> While going into less technical detail, the blog post was much more explicit about justifying methodological claims about the appropriateness of the methods to answer the research question and the benefits of the tools chosen for analysis ('Q9. And what are the benefits of doing this type of analysis?') (<a href="https://www.lshtm.ac.uk/research/centres/march-centre/news/217376/how-does-it-work-innovating-family-planning-interventions">https://www.lshtm.ac.uk/research/centres/march-centre/news/217376/how-does-it-work-innovating-family-planning-interventions</a>)</p>

<b>CEDIL project and lead</b>	<b>What are claims about?</b>	<b>How is the confidence in the claim communicated for different audiences?</b>	<b>Can confidence in the claim communicated be traced back to the original assessments?</b>
<b>Evidence synthesis:</b> 'Advancing evaluation of gender and social equality outcomes in WASH' (Macura)	Relevance of findings	'The review scope is co-designed with stakeholders ensuring relevance of findings for WASH policy and practice.'	Explicit justification in link to blog post. Feedback on the protocol is publicly available at: <a href="https://www.sei.org/wp-content/uploads/2020/05/response-to-feedback-from-open-consultation-1.pdf">https://www.sei.org/wp-content/uploads/2020/05/response-to-feedback-from-open-consultation-1.pdf</a>
<b>Evidence synthesis:</b> 'Technology-Based Innovative Solutions for Improving Perinatal Care Utilization: A Network Meta-Analysis' (Rahman)		No other outputs available as no longer supported by CEDIL due to UK aid cuts during COVID-19.	No other outputs available as no longer supported by CEDIL due to UK aid cuts during COVID-19.
<b>Evaluation:</b> 'Impact Evaluation of Ethiopia's SHARPE programme' (de Brauw, 2022)	Appropriate combination of methods	<b>Blog post:</b> Two specific intervention constraints, and how the intervention may be amended to overcome constraints, were described in plain language. However, the role of RCTs and qualitative research were not described. ( <a href="https://cedilprogramme.org/blog/community-referral-financial-access-refugee-populations-ethiopia/">https://cedilprogramme.org/blog/community-referral-financial-access-refugee-populations-ethiopia/</a> )	There is a link in the blog post to the study home page on CEDIL's website.

<b>CEDIL project and lead</b>	<b>What are claims about?</b>	<b>How is the confidence in the claim communicated for different audiences?</b>	<b>Can confidence in the claim communicated be traced back to the original assessments?</b>
<p><b>Evaluation:</b> ‘Impact of Maternal Depression Treatment on Maternal Health, Parental Investment, and Child Development’ (Maselko <i>et al.</i>, 2022)</p>	<p>Impact of context on research in progress</p>	<p><b>Blog post:</b> Talked about how Covid-19 affected field research in general, and specifically addressing this study. The post was written in accessible language, explained concepts and implications in detail, and did not use jargon. A different audience from the design paper, which was more technical (<a href="https://www.bachpanstudy.com/post/covid-in-pakistan-its-effects-on-field-research-work-and-future-implications">https://www.bachpanstudy.com/post/covid-in-pakistan-its-effects-on-field-research-work-and-future-implications</a>)</p>	<p><b>Blog post:</b> Highlights the potential limitations brought about by COVID-19. The design paper stated that discussion surrounding the impact of COVID-19 on the study was in the accompanying inception paper. (<a href="https://www.bachpanstudy.com/post/covid-in-pakistan-its-effects-on-field-research-work-and-future-implications">https://www.bachpanstudy.com/post/covid-in-pakistan-its-effects-on-field-research-work-and-future-implications</a>), another blog post available is of little relevance to the study</p>
<p><b>Programme of work 2: Enhancing evidence transferability</b></p>			
<p><b>Evidence portal:</b> The Disability Evidence Portal (Shakespeare) <a href="https://disabilityevidence.org">https://disabilityevidence.org</a></p>	<p>Relevance of findings (to disabled people and LMICs)</p>	<p>The claim of relevance applied to evidence accessible through a portal was justified by assessing the degree to which the evidence assesses stakeholder involvement, local relevance, and feasibility in a variety of resource settings.</p>	<p>Explicit justification offered by a publicly available tool.</p>

CEDIL project and lead	What are claims about?	How is the confidence in the claim communicated for different audiences?	Can confidence in the claim communicated be traced back to the original assessments?
<p><b>Evaluation:</b> ‘Machine learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programs in Uganda (7-year follow- up)’ (Chioda and Gertler, 2022)</p>		<p><b>Working paper:</b> Adopted largely the same technical language as it was targeted at the same type of audience as the design paper and not designated for the general public (<a href="https://escholarship.org/content/qt7k99c8vj/qt7k99c8vj.pdf">https://escholarship.org/content/qt7k99c8vj/qt7k99c8vj.pdf</a>)</p> <p><b>Presentation:</b> Accompanied the working paper (<a href="https://www.dropbox.com/s/hr9gvy3pjsls7pc/SEED%20-%20USC%20%2809.25.2021%29%20v5.2.pdf?dl=0">https://www.dropbox.com/s/hr9gvy3pjsls7pc/SEED%20-%20USC%20%2809.25.2021%29%20v5.2.pdf?dl=0</a>)</p>	<p><b>Working paper:</b> Adopted largely the same methodological claims and justifications as the design paper – talked about the study design of the RCT (but only included discussion about one of the interventions (SEED)). (<a href="https://escholarship.org/content/qt7k99c8vj/qt7k99c8vj.pdf">https://escholarship.org/content/qt7k99c8vj/qt7k99c8vj.pdf</a>)</p> <p><b>Presentation:</b> Accompanied the working paper. (<a href="https://www.dropbox.com/s/hr9gvy3pjsls7pc/SEED%20-%20USC%20%2809.25.2021%29%20v5.2.pdf?dl=0">https://www.dropbox.com/s/hr9gvy3pjsls7pc/SEED%20-%20USC%20%2809.25.2021%29%20v5.2.pdf?dl=0</a>)</p>
<p><b>Evaluation, secondary data analysis:</b> ‘POInT Research Design Paper’ (Davey <i>et al.</i>, 2022)</p>	<p>Theory of change and transferability of findings</p>	<p><b>Design paper:</b> They laid out a plan for the diffusion of their results, which included papers, but also blog series, webinars, online tools, guidance, a workshop and videos. How will the presented claims differ in the different diffusion channels? At the moment, no other outputs are available.</p>	

## Appendix 6

### Disability Evidence Portal – Rating System Appraisal Checklist

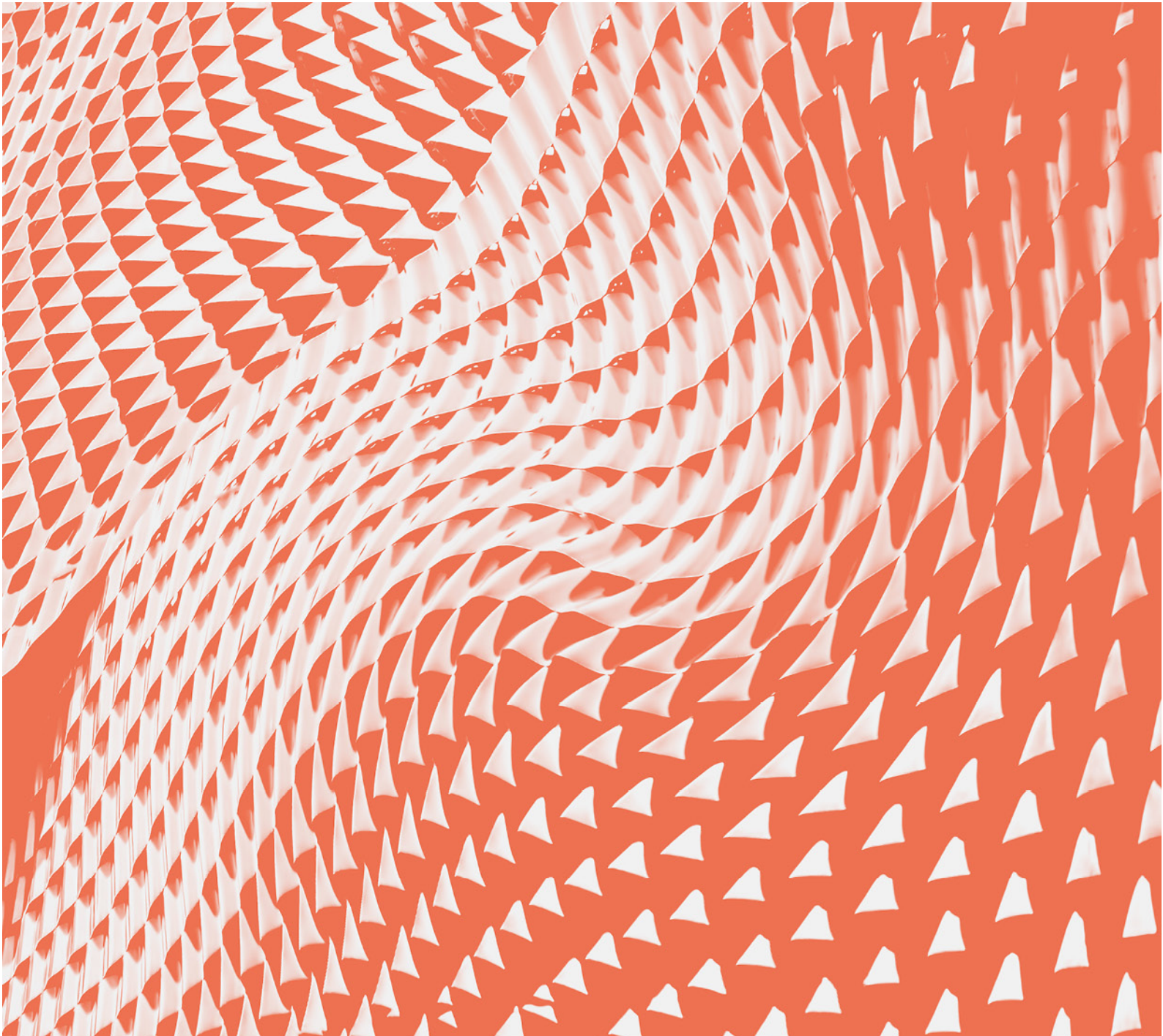
This checklist seeks to inform the process of appraising the quality of research and evidence used in the synthesis of evidence briefs for the Disability Evidence Portal.

Rating Filter #	Description	Criteria for 1 point	Criteria for 2 points	Criteria for 3 points	Author Considerations	Final Score
Equity	The level of active engagement or participation of people with disabilities at various stages of research projects and programmes (design, development, implementation, evaluation) synthesised in this evidence brief	None of the reviews included considerations for minorities or consulted with Disabled Peoples Organisations in the delivery of the project.	Under 50% of the reviews included considerations for minorities OR consulted with Disabled Peoples Organisations in the delivery of the project.	Over 50% of the reviews included considerations for minorities OR consulted with Disabled Peoples Organisations in the delivery of the project.		
Depth	The level of weight, accuracy and depth of literature covering the specific topic of this evidence brief	Under 50% of reviews specific to the topic covered by the evidence brief. High level of	Between 50-70% of reviews specific to the topic covered by the evidence brief. Medium to	Over 70% of reviews specific to the topic covered by the evidence brief. Low level of		

CEDIL Lessons Learned Paper 2: Evidence claims for informing decisions relating to socio-economic development

		bias as reported in discussion.	low level of bias as reported in the discussion	bias as reported in the discussion		
Local relevance	The level to which the literature included in the synthesis of this evidence brief is locally relevant e.g. interventions developed in specific contexts	Majority of the literature coming from HICs (LMIC <20%)	A varied proportion of literature from HIC and LMIC (LMIC 20-50% )	Majority of literature coming from LMICs (LMICs >50%)		
Feasibility	The level to which the evidence and recommendations synthesised can be generalised to the resources available in specific contexts	Could be delivered in a high-resource setting	Could be delivered in high to middle-resource settings	Could be delivered in a wide variety of resource settings	To be discussed with by an advisory group.	





@CEDILProgramme



Centre of Excellence for Development Impact



## About CEDIL

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

For more information on CEDIL, contact us at [cedil@opml.co.uk](mailto:cedil@opml.co.uk) or visit our website [www.cedilprogramme.org](http://www.cedilprogramme.org)