Approaches to Evidence Synthesis in International Development

Sandy Oliver[1], David Gough [1], James Copestake[2]

August 2017

[1] Evidence for Policy and Practice, Information and Coordination Centre, University College London
[2] Department of Social and Policy Sciences, University of Bath

Colophon

Suggested Citation: Oliver S, Gough D, Copestake J, 2017 Approaches to Evidence Synthesis in International Development. CEDIL Pre-Inception Paper: London

About CEDIL: The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium initiative supported by UKAID through DFID. The objective of the centre is to commission and implement impact evaluations, promote the uptake and use of evidence from impact evaluations, and develop and demonstrate new and innovative methodologies for impact evaluation and evidence accumulation. Evidence for Policy and Practice, Information and Coordination Centre (Eppi Cente) is one of a number of CEDIL members that form its intellectual leadership team.

Corresponding Author: Prof Sandy Oliver, email: sandy.oliver@ucl.ac.uk

Table of Contents

Abstract

This paper discusses the spectrum of synthesis methods available to generate, explore and test theory, their value to the field of international development and innovations required to make better use of the primary research available. It goes further than setting substantive priorities for international development impact and learning. It addresses current advances and priority gaps in the methods for research before considering the substantive and methodological direction of evidence synthesis for impact evaluation in international development, particularly as this relates to the work of CEDIL.

This scope encompasses methods for all stages in the process, from setting the question to appraising and synthesising the findings. It describes existing methods for synthesis, including how methods vary and the guidance and standards available. It then considers how well existing methods match the field of international development and the latest innovations emerging or required before providing a research agenda for advancing synthesis methods.

In particular, it argues for clearer distinctions between syntheses produced as public goods, and those tailored to specific circumstances; and strengthening knowledge systems through greater use of maps to navigate existing and missing evidence, harmonised outcomes and measures, and advances in automation technologies. Improved methods and guidance are required for synthesising formative research and investigating contextual factors. Engaging stakeholders and working across academic disciplines support the production of policy-relevant syntheses and inspire methods development.

Section 1

Introduction (Scope, Purpose, Audience and Approach)

The Centre for Excellence in Development Impact and Learning (CEDIL) was commissioned 'to innovate in the field of impact evaluation; design, commission and implement impact evaluations; and promote the uptake and use of evidence from impact evaluations' (DFID 2016). Innovation is anticipated both at the level of primary research and the level of research synthesis. While primary research addresses questions by systematically seeking and analysing new data from the field, research synthesis addresses research questions by systematically seeking and analysing data from existing studies (Gough et al 2017).

This paper was prepared during the early stages of setting up CEDIL. It considers the substantive and methodological direction of evidence synthesis for impact evaluation in international development, particularly relating to the work of CEDIL. The paper has been produced for those readers whose role is to develop or apply state of the art evaluation and synthesis methods, whether as part of commissioning or producing impact evaluations, or wider learning for development. Beyond this readership are the people the research agenda will serve: potential users wishing to understand better the issues they face in their professional lives, needing to choose between alternative policies and practices, or with responsibilities to implement the policies and practices selected.

The paper aims to stimulate discussion among professional specialists involved in evidence synthesis about how evidence synthesis can be performed better. Innovation may be viewed from the supply side: for example, taking technical methods established in one field and adopting them or adapting them in another; or as taking advantage of opportunities arising from new advances, for example, digitisation and data science. But it may also be viewed from the perspective of actual and potential demand, including as a response to the way policy-making and development practice is adapting to an increasingly dense but quality indifferent data environment. Going further, the challenge of improving evidence synthesis is both technical and political, and one that entails addressing supply and demand considerations simultaneously. Given that assumptions and language differ between potential producers and users of evidence, concepts and definitions need to be clarified before considering the priorities for innovations in methods, guidance or support.

Our interest is in bringing together the results of individual studies by aggregating and/or configuring them into an overall answer to the systematic review question. This activity, from the point of setting a question and seeking studies, to appraising and integrating the findings, is sometimes called *systematic reviewing* (particularly when collating quantitative studies), and sometimes called *research synthesis* or *evidence synthesis* (particularly when collating qualitative studies). It includes describing what

studies have been conducted (mapping the research terrain), not only synthesising their findings. The term *synthesis* is also applied to the specific stage of collating the findings of studies included in a systematic review. This paper considers methods for all stages, from setting the question to appraising and synthesising the findings. It defines systematic reviews broadly as reviews of existing research using explicit, accountable and rigorous research methods (Gough et al 2017).

Evidence synthesis informs us about what is known from research, making it fundamental for informing policy decisions about development and, in the words of CEDIL's mission, for promoting 'the uptake and use of evidence from impact evaluations'. The argument is beguilingly simple. Drawing upon multiple studies (rather than cherry picking individual studies) increases the scale (and sometimes the statistical power) and representativeness of the evidence about impact or how impact is realised or failed. Multiple studies also provide an opportunity to appraise the trustworthiness of the evidence and takes into account contextual factors and values apparent in studies from different locations and standpoints. An international example is the question of the optimum duration of exclusive breastfeeding (see Box 1).

Box 1: Evidence underpinning WHO guidelines on breastfeeding (Kramer and Kakuma, 2002)

A systematic review of existing primary research identified:

2 RCTs from Honduras

7 observational studies from developing countries

9 observational studies from developed countries

Overall synthesis of findings:

Exclusive breastfeeding for 6 months, compared with 3-4 months, associated with:

- No apparent growth deficits
- No apparent association with allergies
- Poorer iron status: RCT in Honduras
- Delayed return of menses and more rapid postpartum weight loss in mothers

Undertaking such systematic reviews provides timely access to the research evidence base that policymakers can use to inform policy development (in conjunction with other forms of evidence and other factors influencing decision making). The second purpose of evidence synthesis is to collate what we reliably know from prior research addressing development impact before embarking on new research.

Evidence synthesis for impact and learning is defined for the purpose of this paper as the systematic and transparent integration of evidence to assess or understand impact from secondary sources in ways that makes it more useful. The reference to 'secondary sources' typically reduces the need for the collection of additional primary data from

intended and unintended beneficiaries (and victims) of development activities (though mixed knowledge reviews of previous and new primary research do exist), while the word 'systematic' rules out literature surveys and reviews that do not follow a rigorous and transparent process of selecting, screening and collating evidence. The term 'evidence synthesis' includes both the systematic reviews of evidence of impact (e.g. experimental study designs) and synthesis of the processes involved in seeking to make an impact (e.g. mixed methods research from case studies or process evaluations). This means it spans approaches to evidence synthesis based on a range of assumptions about truth and how to get at it.

One of the most basic frameworks through which to begin to explore the evidence synthesis space is based on a spectrum of research problems or questions from relatively 'simple', through 'complicated' to 'complex'. References to complexity have become increasingly common in the field of international development, where choices between tackling simple-but-superficial and profound-but-wicked problems abound: a problem highlighted by Natsio's (2010) proposed development dilemma that 'the more measurable the action, the less likely it is to be transformative'. Related to this is the so-called "evidence paradox" - or existence of a trade-off between addressing questions rigorously and narrowly, or more broadly but with less confidence (Vaessen, 2016:175).

Impact evaluations tend to aggregate data to measure the extent of impact. Both primary and secondary research and evidence synthesis – particularly aggregative research – is simpler for problems that (a) are widely understood and consistently specified, (b) involve relatively homogenous interventions or treatments, (c) are affected by fewer and/or easily measurable confounding variables, and (d) have effects that are relatively easily identified, measured and located over time (Woolcock, 2013). Complicated problems exist within systems with dense but measurable causal interconnections and clear boundaries. Complex problems, on the other hand, include less predictable interactions and emergent properties, opening up more space for researchers and practitioners to frame them in different and unavoidably partial ways[1]. Systematic reviewers have the option of recognising the inherent complexity in all interventions and their application or, conversely, asking and addressing simple questions about complex problems (Petticrew et al 2015). For instance, the complex problem of daycare may be investigated from the simple perspective of questions about effects that are answered by randomised controlled trials; or it may be investigated by more complex questions to identify 'the components of daycare, perhaps creating typologies of daycare, analysing the barriers and facilitators of effective care, and how different elements in the system interact' and more (Petticrew et al 2015).

The potential for disjuncture between and among suppliers and users of evidence is high. Users may have complex questions but seek, or be offered answers from simple syntheses of evidence. Similarly claims to universal applicability or objectivity about the effect of a particular intervention or 'treatment' based on methods borrowed from the

[1] Pawson (2013) offers a fuller account of complexity, starting with the acronym VICTORE (volition, implementation, contexts, time, outcomes, rivalry and emergence). See also Boulton (2015).

natural sciences are often undermined by simplifications about the complexity of causal process and the contexts in which they occur that are built into the framing of the problem in the first place (Flyvbjerg, 2001). There are arguments, and some methodological support, for avoiding an emphasis on internal validity (encouraged in the natural sciences). Taking evidence from experimental designs addressing development economics (including microcredit, migration, and education interventions) from one context to apply in another may be less justifiable than taking evidence from observational studies to apply in the same context (Pritchett and Sandefur 2015). For instance, a narrow framing of genetically modified seeds as a technical input into the production of particular crops (Klumper and Qaim 2014) avoids addressing concerns about developing context-specific means to promoting monopoly of supply and weakening overall agro-system resilience (Fischer and Ekener-Petersen 2015). Similarly, an inadequate socio-economic framing of effectiveness has weakened much analysis of microcredit impact by ignoring contextual factors affecting programme design or effects, such as social determinants of health including religious attitudes to financial loans and variation in livelihoods (e.g. micro-loans for tailoring rather than more seasonal agriculture). These were revealed by a systematic review of the qualitative research (Peters et al 2016). As with all research, it is crucial to be transparent about what method of research has been used and the limits of the evidence claim that can be made from it.

This discussion highlights how securing agreement over which research issues and questions to prioritise is no less important (and political) for evidence synthesis than for impact assessment at the primary level. Furthermore, on the basis of Keynes' oft-quoted quip that 'it is better to be roughly right than precisely wrong', the task of prioritising issues and questions for review should be carefully distinguished from the question of ease of evaluation or 'evaluability' (Davies 2013). In other words, evidence synthesis effort should focus on what is agreed to be important, rather than dictated by what it is perceived to be methodologically easier to do well. This may entail sometimes employing methods considered by some as less rigorous in order to avoid misleading conclusions about the direction of effect or precision, the latter being described by Manski (2013) as "incredulous certitude." More importantly, it entails maintaining and strengthening a portfolio of evidence synthesis methods capable of responding to any selected problem along the spectrum, from simple to complex, with a minimum loss of credibility[2]. This is not an argument for a lack of rigour but for using the most fit-for-purpose rigorous, transparent and accountable method, rather than a more rigorous but inappropriate method.

The case for adopting a pluralist approach to evidence synthesis for international development can be extended to include advocacy of mixed methods (Harden and Thomas, 2010). More specifically, it is useful to identify steps within any evidence synthesis project that are best tackled using different approaches. For example, a broad

[2] In favouring this approach to thinking about evidence synthesis we do not underestimate the practical difficulties that classification of the complexity of a problem presents. Indeed, much of this paper is taken up with elaborating on how to do so and with what methodological implications. It is also not lost on us that this classification task itself may also be influenced by political judgements (Grint, 2005).

and interpretive narrative synthesis of evolving approaches to a problem may point towards scope for more narrowly framed systematic review, which may focus on whether and how observed causal effects operate taking into account their context (see Box 2 for an example of a two-stage review).

Box 2: A two-stage review (Obuku et al, 2017)

A review addressing a broad and open question of what works for engaging non-state providers in post-conflict and fragile states in primary healthcare service delivery began by mapping the available evidence. This first stage informed the choice of ten specific interventions which formed the focus of the second stage in-depth review: the effects of government and management (contracting out health services, social franchising, public-private partnerships, community empowerment, accreditation/regulation); health financing arrangements (community health insurance, private health insurance, pay for performance, microcredit schemes) and engaging the informal sector (training for traditional birth attendants).

402 studies met the following eligibility criteria to be included in a map of evidence:

1. Setting: Primary health-care facility or community

2. Geographical location: Fragile and post–conflict states

3. Actor: Deliver an intervention by a non-state actor

4. Intervention: Aims to improve selected primary care outcomes

5. Population: General or specific targeted vulnerable populations of the post-conflict state

6. Language: English and other languages

7. Year: Published 1990 – 2015

Most of these studies employed quantitative designs to assess impact (64%), with fewer studies using qualitative approaches, such as case studies or descriptive accounts of non-state actors providing primary healthcare in post-conflict states. Only a small number of studies employed a mixed methods design (7%).

Of these 402 studies, 107 studies met the eligibility criteria for an in-depth synthesis focus on the effects of ten specified interventions:

8. Aim of study: To assess impact of non-state actors in delivery of primary healthcare

9. Study design: Quantitative primary empirical studies with a comparison group or time period.

Evidence synthesis can support policy-makers at different stages in the policy process by asking different questions of the research and thus drawing on different sets of studies (Lavis 2009). Learning about the nature and scale of problems comes from qualitative and observational studies. Assessing potential policy options may draw on effectiveness studies, economics studies, and studies of views and experiences.

Evidence addressing service delivery can come from effectiveness studies of implementation, acceptability studies and process evaluations. These studies include counterfactual designs, narrative research, case studies, ethnography and more. Each approach needs appropriate methods for assessing their quality, synthesizing their findings, and taking into account the context in which they were conducted. A variety of synthesis methods are readily available for such work (sections 2.1 – 2.5) presenting commissioners and researchers with the task of choosing appropriately between them (section 2.6) and the motivation to develop methods where they are currently inadequate for international development.

Pressing research questions for international development are apparent from:

- A priority setting exercise that consulted academics, think tanks, non-governmental and multilateral organizations and resulted in priority questions across 11 themes: conflict and fragility; education; energy; environmental sustainability; food security; governance; growth and employment; health; inequalities; population dynamics; water and development policies, practices and institutions (The Sheffield Institute for International Development 2015)
- UK Aid's four strategic objectives (HM Treasury and DFID 2015): strengthening global peace, security and governance; strengthening resilience and response to crises; promoting global prosperity; tackling extreme poverty, helping the world's most vulnerable and delivering value for money
- Sustainable Development Goals across 17 themes (http://www.un.org/sustainabledevelopment/)
- Priorities of the 2017 G20 summit: 'Urgent questions of our time. The challenges have increased in recent years. Geopolitical conflicts, terrorism, migration and refugee flows, poverty, hunger, and epidemics and increasing climate change place great burdens on societies worldwide and have far-reaching effects on economic development.' (G20 Germany 2017)
- DFID's research review (DFID 2016), with its vision for the production and use of high quality evidence to achieve: impact; scaling up of technological solutions for use particularly by those who need them most; reduction in inequalities, particularly gender inequality; and working with 'the toughest challenges in the hardest places'.

Such priorities require synthesis methods that allow: collecting information on the nature and extent of different phenomena and how they are perceived and understood by different stakeholders; analysis of complexity and contextual influences (for understanding causal pathways with their mediators and moderators, drawing on evidence for use in another context, adapting and packaging interventions), timescales (for resilience and sustainability), and health and social inequalities; the development of policies and practices and their assessment; synthesis including areas where rigorous research is sparse and where circumstances are fast changing; and synthesis of knowledge across and beyond academic disciplines.

Achieving clarity about what is important to investigate, and seeking to advance evaluation methods in international development, are both part of a broader endeavour to add value and avoid waste in research through the prioritisation of the research endeavor and through the use of systematic reviews to clarify what we do and do not know from research (Chalmers et al 2014; Ioannidis et al 2014).

This paper goes further than setting substantive priorities for international development impact and learning. It addresses current advances and priority gaps in the methods for research. Section 2 describes existing methods for synthesis, including how methods vary and the guidance and standards available. Section 3 considers how well existing methods match the field of international development and the latest innovations emerging or required. Section 4 provides a research agenda for advancing synthesis methods.

Section 2

Existing Methods for Synthesis

Bearing in mind our broad definition of systematic reviews, evidence synthesis for impact and learning as the *systematic* and *transparent* integration of evidence to assess or understand impact from secondary sources in ways that make it more useful, all evidence syntheses follow a transparent process of selecting, screening and collating evidence. The particular system or method depends on the purpose and question of each synthesis. This section considers a spectrum of methods available.

2.1 How Synthesis Methods Vary

Methods of systematic review are similar to primary research methods. This is not surprising as systematic reviews are simply another level of research analysis. Primary research aims to answer research questions by engaging with, collecting and analysing data from the field. Reviews of research ask the same questions but instead of engaging directly with the world and collecting new data, they engage with the findings of studies that have already been undertaken. They are simply a secondary (meta) level of analysis.

Primary research studies vary in their general approach and design depending on the questions they address. For example, a question asking about the extent of phenomena (prevalence) calls for an approach and design that is likely to be different from a study examining the extent of the impact of an intervention or the process by which an intervention has an impact. These differences in approach and design are likely to be reflected in evidence syntheses in two ways.

First, a review asking a particular question is likely to include in the review the type of primary studies typically used to answer that research question. So, reviews asking questions of process will, of course, include primary studies on process, whilst reviews

addressing questions of impact will include counterfactual evaluations such as controlled experimental studies of impact or analyses that use matching and/or statistical controls.

Second, the general approach and design of the review is likely to reflect that of the primary studies. So, reviews that ask questions of causal process are likely to reflect the approach and design of the primary process studies, by configuring findings from such studies, whilst reviews that ask questions of impact are likely to use review methods that adopt the same design features and methodological assumptions as counterfactual evaluations, namely aggregating data.

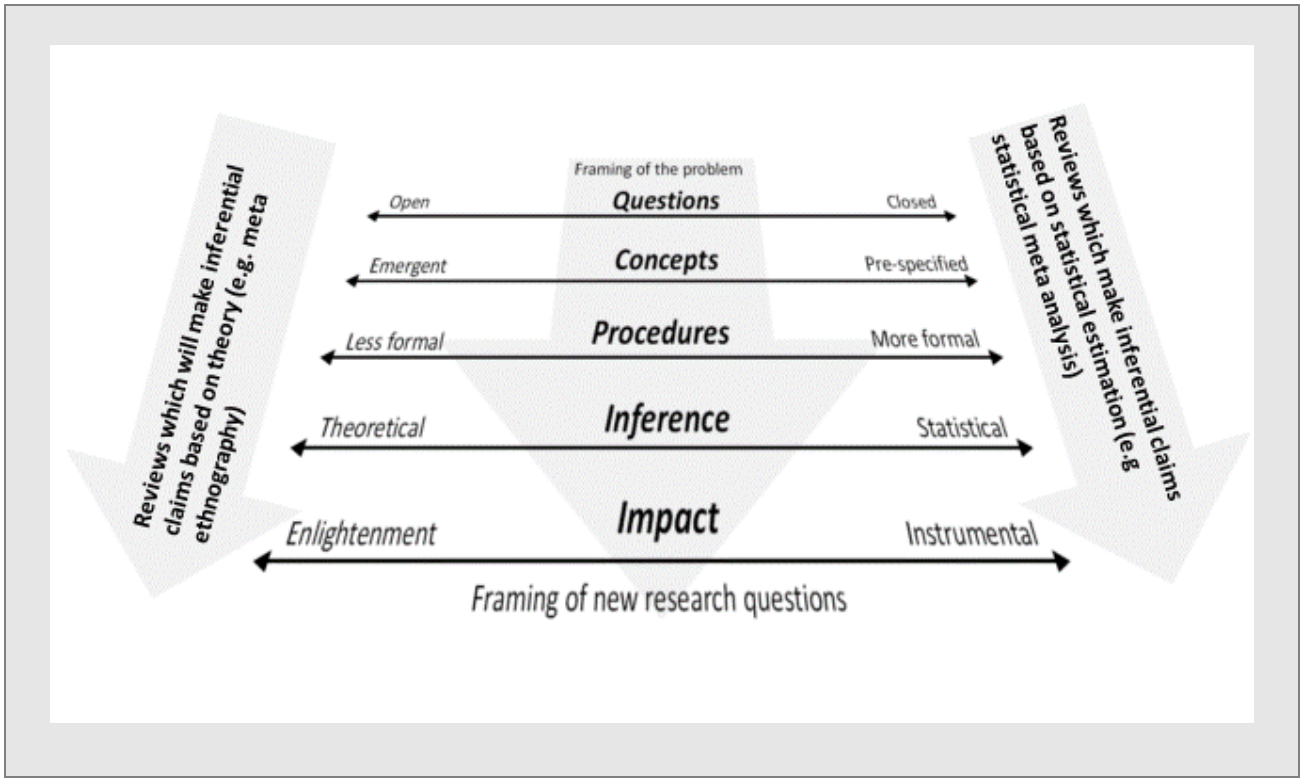
Because of these similarities between primary research and reviews of that research, the explanation below of variation in review methods considers many of the standard issues in research methods and design including the research question, the general approach and design, the relationship of the research question to theory, the extent that aggregation and configuration are used in the analysis of data, the extent of a research problem addressed by the study, and components and stages of the study. Reviews, just like primary research, vary on many dimensions (that we call 'dimensions of difference') (Gough et al 2017). This variation in approaches, design and dimensions of difference in method leads to distinct types of review methods, some of which are named and widely recognised. After considering these variations in the rest of section 2, we consider in section 3 how to choose methods to suit the needs of international development work. As with all research, appropriate, relevant, rigorous and transparent methods need to be applied in order for the research to be considered trustworthy and able to provide justifiable evidence claims (Liabo et al 2017). Traditional literature reviews may not meet the required evidence standards in order to make evidence-based, justifiable claims. A common problem is lack of transparency in the reporting of the method which limits the verifiability and thus accountability of the method.

2.1.1 Type of Question, Relationship to Theory and Research Design

Systematic reviews, like all scientific research, use explicit rigorous methods, although these methods can vary in the degree that they attempt to generate, explore or test theory. In some cases, the aim will be to develop or explore new ideas, concepts and theories which is likely to require an iterative investigative inductive approach. In other cases, a theory will be the basis for collecting information on the assumptions of that theory. In further cases, the collection of such data may be used to collect data to empirically test a pre-specified theory.

Research questions that are developing and exploring theory are likely to have open questions, emergent concepts, to have more iterative and less formal procedures, to use theoretical inference and to provide insight and enlightenment. In contrast, research questions that are testing theory are more likely to have closed questions with pre-specified concepts (the prior theory), to have more a priori defined formal procedures, to use statistical inference and to create data and 'facts' to be used to inform decisions (see Figure 1).

Figure 1: 'Dimensions of difference' in approaches to systematic reviews (Figure 3.11 in Gough et al, 2017)



2.1.2 Aggregating and Configuring Research Analysis

Research also differs to the extent that the data analysis is adding up or arranging and re-organising data. This distinction is a bit similar to that between quantitative and qualitative analysis. Adding up is quantitative and arranging is qualitative. However, the terms quantitative and qualitative are not very precise as they are referring to data types rather than the approach to analysis. You can aggregate data even if it is not in quantitative form and you can configure qualitative data.

The aggregation/configuration distinction relates to the dimensions of difference in approach in Figure 1. The testing questions are more likely to have specified theory, formal methods and to aggregate data. In a primary experimental controlled study or a systematic review of such studies, there will be a priori methods to test pre-specified hypotheses but there may also be a post hoc iterative configurative exploration of the data that generate new hypotheses for future experimental testing. Similarly, a theory generating primary study or review may also aggregate some data as part of the exploratory investigation.

Within this spectrum of approaches, a number of methods have evolved to address questions that relate to theory in different ways.

Testing theory: Many systematic reviews test hypotheses about the effects of an intervention or exposure. Some of these systematic reviews include statistical meta-analysis:

- Statistical meta-analysis: a priori approach to the synthesis of primary studies to test the impact of interventions. Such reviews include studies with control groups, either constructed before introducing an intervention (such as controlled trials) or constructed by matching groups (such as cohort studies) or accounting for differences between groups statistically (such as with multivariate or instrumental variables regression).
- Network meta-analysis: a development of meta-analysis to compare the relative impacts of different interventions. The different studies considered (in the network) may not have individually compared the impact of different interventions but examining the studies together enables the assessment of comparative effect. Such reviews include studies with control groups (such as randomized controlled studies with overlapping treatment and outcome factors).
- Decision modelling: developing predictive models (including Bayesian models) of how variables interact, to lead to different outcomes (allowing many types of data to be included). Such reviews include various experimental, large-scale observational and case studies (Sainfort et al 2013).
- Predictive test accuracy: reviews that aggregate the evidence on the extent that a predictive test is predictive. Such reviews include cross-sectional or case-control diagnostic test accuracy studies.

Generating theory: In contrast, other methods develop new theory:

- Meta-ethnography: ethnography of qualitative research to develop concepts/theories about the nature of the phenomena under study (Toye et al 2014) (such as ethnographic or similar inductive qualitative primary studies)
- Thematic synthesis: 'bottom-up' synthesis to inductively develop concepts from themes identified in primary studies (less dependent on ethnographic theory) (Thomas and Harden 2008) including any type of data where themes can be identified.
- Meta-narrative synthesis: similar to framework synthesis (see next sub-section) but focuses on the historical development of concepts within an area of study (Greenhalgh et al 2005) including any studies that describe concepts used at any point in the historical development of an idea or approach.

Exploring theory: Between these two extremes are methods that explore existing theory:

- Framework synthesis: thematic synthesis that is 'framed' by different degrees of a priori and iterative development (Dixon-Woods 2011) including any study with data that can be coded and interrogated.
- Qualitative Comparative Analysis: for analysing small numbers of studies (where statistical analysis can be limited), and explaining a given outcome by identifying multiple pathways of putative causal factors (Candy et al 2013, Brunton et al 2014) including any statistical studies of causal effect.

Mixed approaches to theory: Some reviews mix their approaches to theory

- Mixed methods reviews: such as a review with two arms that examine two sub-questions, one of which maybe testing hypotheses and one which may be exploring concepts.
- Realist synthesis: two-stage review that configures the theory of change within a hypothesis of cause and then aggregates the evidence to test each part of the causal model (though it takes an iterative approach to such testing), including many designs of primary studies.

Non-theory specific approaches: That can be applied to theory generating, exploring or testing reviews

- Prevalence synthesis: the aggregation of prevalence data from multiple primary prevalence studies including surveys and case-controlled studies.
- Economic costs: synthesis of the evidence of costs of phenomena. This can be modelled or combined with hypothesis testing reviews of the impact of interventions to assess cost and impact together, including many types of data on costs.

It is important to emphasise that these differences between reviews are relative dimensions and not absolute distinctions. Even the most a priori study is likely to have some iteration and exploratory studies may have some conceptual assumptions. Framework approaches in both primary research and reviews may be predominantly exploratory and configuring, yet be driven by some pre-specified concepts that frame the method of data collection and analysis.

2.1.3 Extent of the Research Problem Addressed and Multi-Methods Approaches

Just as with primary research, reviews vary in their ambition and how much they attempt to achieve. They vary in the breadth of the question addressed and the depth in which they answer it. Reviews with more finance, staff, and time resources can attempt to achieve more than those less well resourced. This is not the same as the extent of the evidence claims that can be made by the findings of a primary study or a review of such studies as this will depend on many other factors such as the nature of the evidence available and quality and relevance of the research methods applied (Gough and Thomas 2017).

Some reviews aim to provide a non-systematic overview of research findings which can be quick and cheap but, as with any non-rigorous piece of research, may provide misleading results. Other reviews may be rigorous and transparent and so systematic but still be rapid by narrowing the breadth, lessening the depth of question addressed, reducing the reliability of the work by searching fewer sources, or minimising coding conducted by two or more reviewers independently (Featherstone et al 2015) (further detail on rapid reviews appears below).

Another approach to covering more ground without doing more work is changing the starting point by systematically reviewing existing systematic reviews. Building on prior

reviews allows: review level analysis that focuses on the findings of each review rather than their included studies (with possible updating and some new synthesis of primary studies for any gaps in coverage across the existing reviews); fast access to primary studies identified by prior reviews, especially for issues that are commonly addressed in full texts but not titles and abstracts; and/or reanalysis of studies for cross-cutting issues such as inequalities or ethics.

Other reviews take an opposite strategy and are extensive in breadth and depth by including several components or stages to address different aspects of a question. For example, a mixed methods review with two sub-questions and two synthesis 'arms', one using a priori methods for testing the hypothesis that an intervention evidence is effective and the other exploring stakeholder views about the nature of, or their experiences of, factors related to the intervention (such as in the mixed methods in the bulleted list above, Thomas and Harden 2008). Another example is a two-stage review that first configures the logic model underpinning a hypothesis that an intervention has a causal effect and then second aggregates the empirical evidence for the different stages of that logic model or theory of change (e.g. realist synthesis or causal chain analysis) (Kneale et al 2015). Such approaches harness the strengths of both the exploration and testing of theory and move us beyond 'black box' hypothesis testing of impact with little understanding of causal processes.

2.1.4 Updating Systematic Reviews

A final issue is how the development of research over time is addressed by reviews. One aspect of this is how the results of synthesis may vary over time. In a statistical meta-analysis to assess the effect size (impact) of an intervention, each time a new primary study is undertaken, the synthesis of all relevant studies is likely to change. It is, therefore, possible to plot how our knowledge of the extent of the effect size changes historically. If there is a clear result that does not differ over time showing for example that the intervention does or does not have effect, then it may not be necessary and thus wasteful to undertake further primary studies.

A related aspect of how reviews are situated in time is provided by living reviews. These are never complete reviews but are always seeking the latest study that meets the inclusion criteria for the review. This may not seem that different from a more general expectation that reviews should be regularly updated to reflect the latest evidence. The difference is that the existence of the living review can change how you approach primary studies. An experimental controlled primary study, for example, would undertake a power calculation to determine what sample size was necessary for the study to be able to identify the impact of an intervention. With a living review being available, the primary study can undertake its power calculation on what sample size would be necessary to change the findings of that living review, for instance, to shift the total effect size from the categories of negative impact, no impact, or a positive impact, or between one of these categories to another, or between small, medium or large effect sizes (Garner et al 2016). Judging whether a statistically significant difference is considered important requires translating the effect size (a metric-free ratio) into a more meaningful policy metric such as the number of food portions in healthy eating (Thomas et al 2004) or pupils' additional month's progress in education (Higgins et al

2016). This approach to power calculations is a subtle but important change in focus from primary studies to reviews. It provides an understanding of what is known from research to date and how new primary research may or may not change that understanding.

2.1.5 Rapid Systematic Reviews

Increasingly common are efforts to produce systematic reviews within tighter time frames or budgets, namely rapid reviews (Featherstone 2015). Accelerating the pace of systematic reviews can occur early when clarifying the question and restricting the focus in discussion with key stakeholders and gauging likely fruitful areas of the literature; and later when conducting the review through appropriate choice of methodology, and taking shortcuts in identifying, describing, appraising and synthesising studies (Thomas et al 2013; Oliver et al in press).

In these circumstances, the first choice is between adopting or adapting evidence already synthesised (Table 1) and synthesising evidence tailored to the specific need. Existing syntheses can be identified rapidly where they have been made available as public goods in specialist libraries, or through electronic searches of bibliographic databases on the World Wide Web. If necessary this evidence can be re-analysed for local need. Some synthesis methods, such as meta-ethnography, critical interpretive synthesis or thematic synthesis, are designed for slow thinking to maximise understanding rather than working fast to meet deadlines for action. Because their starting point is the minute details reported in manuscripts being reviewed they require extended time for collective deliberation by the review team. In contrast, early application of a framework to synthesise the data can accelerate the process of extracting data and preparing tables. The initial framework is either constructed by the review team taking into account the prior knowledge, priorities and constraints of stakeholders (Oliver et al 2008) or 'borrowed' from theory in comparable fields (Booth and Carroll 2015). Working with static frameworks facilitates fast teamwork but may limit additional learning that arises from frameworks evolving in response to concepts emerging from the literature during an iterative, but slower, review process (Oliver et al, in press).

Empirical evidence about the value and limitations of rapid reviews is growing (Ganann et al 2010) but how well this evidence applies to international development reviews is less clear.

Table 1: Adapting Existing Syntheses for Urgent Evidence Needs

Design of synthesis	Examples
Systematic reviews of systematic reviews: (Overviews) summarise existing evidence, frequently without further statistical analysis	An analysis of inequity arising from public health interventions included only systematic reviews that discussed differential health effects by socioeconomic status (Humphreys and Ogilvie 2013). A series of rapid systematic overviews of systematic reviews of randomised controlled trials (RCTs) of self-management support interventions ('quantitative meta-

	reviews'); and systematic overviews of systematic reviews of qualitative studies of patients' experiences relating to self-management ('qualitative meta-reviews') revealed core components of effective support for patient self-management (Taylor et al 2014).
Updates of systematic reviews: Existing systematic reviews supplemented by updating the literature searches.	A rapid review to inform ongoing UK Department of Health decision-making on how to increase rates of generic prescribing located 10 reviews, gleaned their individual studies and additional studies from "top-up" searches to cover intervention types that were not adequately addressed in the original reviews before the final analysis (Moe-Byrne et al 2014).
A multilevel synthesis strategy combining the above methods: Provide reliable knowledge to answer multiple related questions.	A rapid evidence synthesis (Paton et al. 2016) to support delivery of emergency mental health treatment sought evidence for key interventions at each point in the intervention pathway, according to a hierarchy of evidence sources: first, relevant, empirically supported guidance; second, overviews of reviews; third, systematic reviews; and fourth (where no relevant evidence synthesis existed and gaps were identified by the service user group), primary studies identified through database searches.

2.2 Guidance and Standards for Producing Systematic Reviews

The various approaches to systematic reviewing are described in guidance offered by organisations that support networks of systematic reviewers. Guidance for conducting reviews typically emphasises the systematic steps of identifying and synthesising the evidence (see, for instance, handbooks from Cochrane and the Joanna Briggs Institute). Some synthesis methods are sufficiently widely adopted to have agreed methodological standards for reporting or appraising systematic reviews: PRISMA (Liberati et al 2009); AMSTAR (Shea et al 2007); MECIR (Higgins et al 2016); RAMASES (Wong et al 2013a, b); ROBIS (Whiting et al).

However, many of these standards and tools address internal validity more than external validity. External validity is often poorly addressed (see Pearson and Coomer (2009) for a case study from the field of substance misuse). As external validity may be particularly important for taking into account contextual factors in reviews for international development, an extension to the PRISMA tool, PRISMA-Equity, was developed to prompt reviewers to take into account some social determinants of health inequalities (Welch et al 2012).

Additional guidance which focuses on developing the conceptual framework is particularly relevant to the broad scopes or complex issues addressed by many international development reviews (Waddington et al 2012; Gough et al 2017). So is guidance for rapid systematic reviews, which is only recently available and as yet

untested for health systems research with an emphasis on developing countries (Tricco et al in press).

Section 3

Applying Evidence Synthesis Methods to International Development

So far, this paper has considered how syntheses vary, and the standards that have been developed, in methodological terms generally. This section moves the focus specifically to synthesis for international development.

3.1 Matching Evidence Synthesis Methods to International Development

In 2010, DFID made a large investment in systematic reviews. A study of early experiences of all the main stakeholder groups engaged with this programme of systematic reviews were gleaned from key documents, around 25 interviews and feedback from over 30 completed questionnaires (Rose and Battock 2012). Headline messages about DFID's methods included the value of:

- devoting effort to identifying and developing suitable systematic review questions;
- maintaining continuity and commitment of all stakeholders throughout the lifetime of a review;
- adopting a two-stage process for systematic reviews that splits scoping and analysis;
- systematic review methodology, notably on how to synthesise qualitative evidence;
- identifying and involving from the outset potential end users of the findings; and
- keeping systematic review findings fully independent of existing DFID policy, to allow orthodoxy to be challenged.

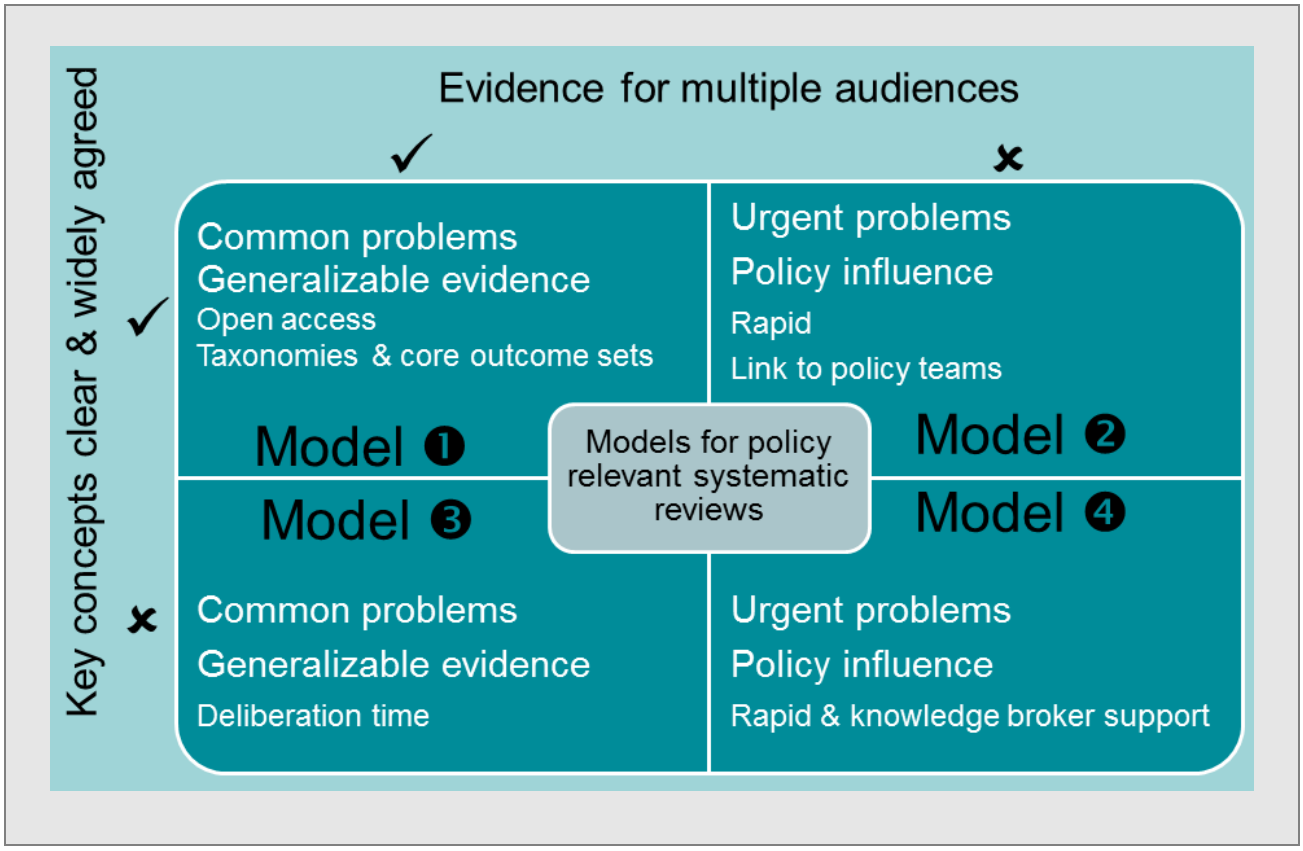
Subsequently Langer and Stewart (2013) researched author reflections of the application of systematic reviews to (mainly effectiveness) questions from international development. Their thematic analysis identified a steep learning curve for reviewers attempting to transfer best review practice from those of health care to international development, and a need to adapt methods to suit international development: multidisciplinary teams to bring sufficient topic knowledge and methodological skills; conducting systematic maps as a first step to assessing the literature available; translating broad questions from policymakers into narrower, manageable questions for synthesis; emphasising programme theory to structure a synthesis; and including a broader range of literature, some of it unpublished, particularly to address programme

theory and contextual factors that may influence impact. Below we consider in detail how systematic reviews vary in their application to international development.

3.1.1 Matching Synthesis Methods and Innovation to the Context of Policy Making

The EPPI-Centre has worked with several policy organisations funding systematic reviews for international development. From reflecting on our experience and interviewing research and policy stakeholders (Oliver and Dickson 2016), we have distinguished systematic reviews in terms of: whether they are produced either as a 'public good' for a general audience because problems are widespread and enduring (models 1 and 3), or as a product tailored to specific concerns in mind; and whether prior consensus on key concepts and definitions is strong or weak (models 2 and 4) (Oliver and Dickson 2016). Figure 2 combines these in a two-by-two matrix.

Figure 2: Four Models of Policy-Relevant Systematic Reviews



Model one is applicable when there is a good consensus about key concepts and definitions for addressing important, common problems. Reviews are made available as 'public goods' in databases such as *The Campbell Library* and *The Cochrane Library*. Model two suits reviews that are needed urgently for specific issues and can make use of clear concepts and definitions. Model three to producing 'public goods' reviews addressing common, enduring problems where key concepts and definitions are not clear or widely agreed in advance. Model four is for reviews tailored to specific circumstances where key concepts and definitions are not clear in advance. Models three and four achieve clarity over key concepts and definitions as part of the review production, either

through widespread consultation and consensus development (typically model three) or with the support of a knowledge broker (typically model four).

The policy relevance of tailored reviews aligned with models two and four is typically achieved through discussion with policymakers initiating (and funding) the review, sometimes with the support of a knowledge broker (Campbell et al 2011; Moore et al 2017). The policy relevance of public goods reviews aligned with models one and three can be achieved through formal priority setting exercises (Viergever et al 2010). These are being adapted for systematic reviews conducted by the Cochrane Collaboration (Nasser et al 2013) and for reviews addressing health systems in low and middle-income countries, where scarce resources make priority setting particularly important (Akl, under peer review).

These models can interact to maximise efficiency in knowledge production and use. When evidence is required urgently, reviews tailored to the specific situation (models 2 and 4) may benefit from drawing on or re-analysing 'public goods' reviews already available (models 1 and 3). Conversely, once new reviews for urgent policy dilemmas have been completed rapidly (models 2 and 4), they may be re-visited and through further discussion and more comprehensive searching and analysis as required, be converted into 'public goods' reviews (models 1 and 3).

Because systematic reviews are relatively new to international development, it has not been unusual for individual systematic reviews or review programmes to be combined with building capacity amongst researchers and policymakers. Novice systematic reviewers require longer to complete individual reviews and need to accrue experience before applying state of the art methods which many complex issues require. Also, the investment in 'public goods' systematic reviews has not surprisingly been driven by policy teams with their own immediate policy concerns. Public goods systematic reviews address widespread problems where key concepts are clear and widely shared, but their production can be a poor match for the timescales and responsibilities of policymakers facing urgent problems. In contrast, are reviews that address immediate concerns of specific policy teams who may require topical, contextualised, rapid reviews that are locally applicable.

Box 3: Research recommendations for distinguishing public goods reviews from tailored reviews

We recommend:

- Research funders invest in producing and cataloguing 'public goods' systematic reviews in an open access global evidence repository (Model 1).

- Decision-making organisations raise awareness and skills for their staff to draw on such a global repository of systematic reviews (Model 1).

- When synthesised evidence is required for urgent decisions, systematic methods be applied (Models 2 and 4), and systematic reviewers draw on these repositories to use or re-analyse existing systematic reviews containing relevant evidence.

- When evidence is synthesised fast (models 2 and 4) owners of systematic review repositories investigate opportunities to transform this evidence into 'public goods' systematic reviews (models 1 and 3).

- Investigating the relationship between public goods reviews and tailored reviews.

3.1.2 Matching Synthesis Methods and Innovation to the Steps in the Policy-Making Process

Systematic reviews can be considered relevant to policy (and policymakers) when they present findings clearly for policy audiences to illuminate policy problems; challenge or develop policy assumptions; *or* offer evidence about the impact or implementation of policy options; *and* take into account diversity of people and contexts (Oliver and Dickson 2016). This breadth of interest may require different synthesis methods to be aligned with steps in the policy process of characterising problems, considering policy options, and implementing policy decisions, as identified by Lavis (2009).

Characterising problems: Learning about the nature and scale of problems from qualitative and observational studies calls for a combination of review methods: configuring research findings from qualitative studies of stakeholders' views and experiences to understand problems better; and aggregating findings from surveys, cohort studies and (less often) administrative databases to assess their scale (see Box 4).

Box 4: Synthesis methods for characterising problems

Understanding failures with treatment: When statistical meta-analysis showed that standard ways of encouraging patients to take their tuberculosis therapy regularly enough and long enough to be effective were not working (Volmink and Garner 2006), qualitative studies were brought together in a synthesis to compare the factors associated with good or poor adherence to conclude which factors influence adherence to tuberculosis treatment (Munro et al 2007). Eight major factors were incorporated into a simple model of adherence to tuberculosis treatment being influenced by four interacting sets of factors—structural factors (including poverty and gender discrimination), social context factors, health service factors, and personal factors (including attitudes towards treatment and illness). This understanding called for the development of patient-centred interventions that take into account poverty and gender, incorporate support systems tailored to individual patients, and share treatment decisions with patients.

The scale of mental health problems: The development of clear criteria for diagnosing mental disorders in the late 1970s led to numerous surveys. The subsequent development of guidelines for conducting and reporting meta-analyses of observational studies in epidemiology (MOOSE) (Stroup et al 2000) and the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) (Liberati et al 2009) allowed such surveys to be synthesised using widely accepted systematic review methods (Steel et al

2014). The high prevalence of common mental disorders globally recognised by this synthesis has informed over 200 subsequent articles addressing various conditions and geographic regions.

Assessing service quality: In a comparison of private and public ambulatory health services in developing countries, a statistical meta-analysis found that both were poor quality in terms of infrastructure, clinical competence and practice, although the private sector performed better in relation to drug supply, responsiveness, and effort. It was the synthesis of qualitative data that revealed the private sector to be more client-centred (Berendes et al 2011).

Assessing potential policy and programme options: Assessing potential policy options can draw on effectiveness studies, economics studies, and studies of views and experiences (Box 5).

Identifying implementation considerations: Evidence addressing service delivery is similar to studies of interventions. Implementation studies can address the need for implementation, the process, extent, and most effective approach to implementation. Evidence thus also comes from effectiveness studies of implementation, acceptability studies and process evaluations. Such a review addressed the factors influencing uptake of cookstoves for cleaner air and accident prevention in homes (Puzzolo et al 2013).

Box 5: Synthesis methods for assessing potential policy and programme options, and their implementation

Rationalising the workforce: A global shortage of healthcare workers prompted the World Health Organisation (WHO) to convene international deliberations about re-allocating tasks from highly qualified cadres to less qualified cadres or lay health workers supported by additional training. These deliberations and subsequent WHO guidelines for task-shifting (WHO 2012) were informed by syntheses comparing the effects of substituting doctors with nurses or other clinicians; training traditional birth attendants; and delivering maternal and child health services with lay health workers. This effectiveness evidence was complemented by syntheses about acceptability, feasibility, safety, implementation and large-scale programmes to consider the challenges of scaling up.

Not only did this exercise require a range of synthesis methods, it also prompted methodological innovation in synthesis. To support the guideline panel, make use of qualitative evidence in their debates and decisions, a tool was developed to assess the confidence that can be placed in reviews of qualitative research, the CERQUal tool (Lewin et al 2015).

Improving indoor air quality: WHO updated and strengthened its guidelines for indoor air quality by drawing on newly commissioned or published, systematic reviews (WHO 2014). These included reviews of household air pollution and exposures (summaries of recent systematic reviews and other types of evidence such as exposure-

response function models), the effects of interventions such as replacing open fires with clean fuel cook stoves on illness and accidents (meta-analysis and narrative synthesis of effectiveness studies), factors influencing the adoption and sustained use of improved cook stoves and clean household energy (thematic and tabular/narrative syntheses), and the financial implications of trying to improve air quality via changes in cooking technologies and fuels (micro and macroeconomic analyses).

Box 6: Recommendations for research guidance

Guidance for research synthesis, whether aimed at producers or users of reviews should clarify:

- the essential principles that underpin all approaches to research synthesis

- how approaches and detailed methods can vary for different purposes

- how to choose appropriate methods depending on the question and literature available.

3.1.3 Transdisciplinary Teams Bringing Topic Knowledge and Methods Skills

The value of a mixed team between them bringing topic expertise and methodological expertise, originally recognised for the conduct of clinical practice systematic reviews, is also seen as important for reviews in international development (Langer and Stewart 2014). The additional challenge in reviews addressing international development is the multidisciplinary nature of this area of study. Multidisciplinary research (disciplines analysed in parallel before discussing their different findings), offers limited learning compared with interdisciplinary learning (analysing disciplinary interfaces) or transdisciplinary research (transcending traditional boundaries to investigate issues in a holistic way) (Choi and Pak 2006). Such research requires skills for working across boundaries to combine knowledge from different academic disciplines and beyond (Choi and Pak 2007). However, the range of disciplines contributing studies to any particular synthesis is not necessarily clear in advance when building a team to conduct the work. Similarly, the principle of involving patients with experience of a specific health condition in reviews of clinical practice is relatively simple compared with involving the relevant stakeholders in reviews that are more sensitive to contextual influence and yet meant to serve worldwide audiences.

Box 7: Research recommendations for transdisciplinary research

Systematic reviewers and their review partners should draw together their collective experience of working across and beyond academic boundaries on an international scale and develop guidance for their peers.

3.1.4 Conducting Systematic Maps to Navigate the Literature

Synthesis products now vary in scale from individual systematic reviews to systematic reviews of reviews that synthesise findings of research, and from systematic maps of the research literature to libraries of evidence (e.g. The Cochrane Library, The Campbell Library, Evidence AID) that facilitate access to evidence. A recent systematic map of systematic maps revealed that the design of existing systematic maps as currently presented has more potential to serve vertical programming (e.g. HIV/AIDS, microfinance) than horizontal programming (e.g. city planning, neighbourhood services). Specific gaps in the mapped evidence include transportation, urban development, economic policy, energy and disaster risk reduction, which are prime examples of horizontal programming (Phillips et al 2017).

Box 8: Research recommendations for evidence gap maps

CEDIL, in discussion with curators of evidence libraries and evidence gap maps, should consider the options available for navigating evidence sources by different potential users. This should be done in discussion with potential users whether they have responsibility for specific programmes, specific populations or specific geographical areas.

3.1.5 Addressing the Breadth and Complexity of Policy Questions

Many systematic reviews frame their analyses in terms of specifying Population, Intervention, Comparison and Outcomes (PICO). This serves well those reviews of interventions for homogeneous populations where the usual practice is fairly uniform and the important potential effects follow the intervention fairly closely. The world of social and environmental interventions is rarely that simple. Moreover, policymakers have been typically asking much broader questions than researchers. Instead of asking whether a specific intervention has an impact on a specific population in terms of specific outcomes, policymakers may ask what interventions work when outcomes of interest are broad, or vaguely unspecified (see Box 9 for examples). Consequently, although PICO may broadly frame many systematic reviews addressing causal relationships, it is inadequate for making sense of the evidence collated to answer broad questions. Such reviews require methods to both configure and aggregate research findings and often accommodate evidence both assessing and explaining impact (or lack of it). Moreover, for reviews that are generating theory about impact, the PICO may be the output of the research rather than the starting point.

Box 9: Examples of broad questions asked by DFID policy teams where review teams configured and aggregated research

What policies and interventions have been strongly associated with reductions in in-country income inequality? (Anderson et al 2016a)

What policies and other interventions have been strongly associated with the translation of growth into reductions in income poverty? (Anderson et al 2016b)

Under what conditions do the following elements of an education system improve system efficiency, service delivery and learning outcomes, especially for the poorest? a) monitoring systems b) inspection systems c) assessment systems (Eddy-Spicer et al)

How effective are different approaches to engaging with non-state providers in improving the delivery of primary healthcare in fragile, conflict or post–conflict settings? What is the impact of non-state actors' delivery of primary healthcare in fragile, conflict or post-conflict settings? (Obuku et al 2017).

What is the evidence on top-down and bottom-up approaches to improving access to water, sanitation and electricity services in low-income or informal settlements? (Annamalai et al 2016).

Systematic reviewers have responded to this challenge by constructing conceptual frameworks that take into account heterogeneity of interventions and populations, or causal pathways/logic models that delineate interim steps between intervention and final outcomes (Kneale et al 2015). This has proven useful for international development reviews (Snilstveit 2012). Current approaches are still relatively simple in terms of the extent that they take account of the potential complexity in causal models and the consequences for mixing methods in primary research (Humphreys and Jacobs 2015). Review methods are similarly simple in how they synthesise such research and identify the necessary future primary research to help further develop causal theory.

In this field, policymakers' international remits prompted the NGO[3], 3ie, to emphasise external validity as well as internal validity in their tool kit, in order to generalise evidence from systematic reviews using a theory-based approach (Waddington et al 2012). Policymakers' broad questions about policy options with long causal chains prompted reviews framed by a programme theory for the intervention and including a mixed-methods approach with a broad range of evidence while maintaining the rigour and transparency that characterise systematic reviews (Snilstveit 2012). Also, a literature offering more natural experiments than controlled trials has prompted reviewers to maximise the learning available from quasi-experimental designs (Duvendack et al 2012). A systematic review is currently underway examining the use of process evaluations of randomised controlled trials of complex interventions (Liu et al 2016). In the meantime, we have an example of a broad systematic review configuring and aggregating evidence along different causal pathways (Box 10).

[3] An international grant-making NGO promoting evidence-informed development policies and programmes

Box 10: Employing simple causal pathways to configure and aggregate mixed approaches to investigating impact (Example from Annamalai et al 2016)

Review question: What is the evidence on top-down and bottom-up approaches in improving access to water, sanitation and electricity services in low-income or informal settlements?

Effectiveness of 'top-down' pathway: Services provided by public or private agencies through centralised planning and implementation appeared effective in individual studies for connecting populations to water, sanitation and electricity. However, where studies were sufficiently similar to justify pooling findings in a statistical meta-analysis, this conclusion was not confirmed.

Effectiveness of 'bottom-up' pathway: Participatory (bottom-up) approaches adopted by NGOs and CBOs suit the construction and maintenance of toilets, which can be standalone, and statistical meta-analysis confirms their effectiveness for individual but not community toilets. Although studies of bottom-up approaches to improving water access appeared positive more often than studies of top-down approaches, this difference was not statistically significant in a meta-analysis.

Qualitative synthesis of contextual factors suggested a need for the customisation of solutions to meet local needs, and better delivery of services by alternative/non-government service providers. Neighbourhoods without the security of tenure were rarely served well top-down. Bottom-up approaches were also limited in this context, and also in Africa, where efforts may be hampered by particularly modest levels of economic development. Public-private partnerships show promise for top-down approaches to improving water supply. Bottom-up, NGO led initiatives for improving water supply need the co-operation and support of the public sector.

3.1.6 Taking into Account Contextual Differences and Inequalities

Scaling up interventions or introducing them into new contexts raises questions about whether interventions developed in one context can be applied in another, and whether evidence of impact developed in one context can be transferred elsewhere and requires methods that take into account contextual differences and inequalities (Wang et al 2005).

Drawing on syntheses of international evidence for decisions about specific locations, or to inform efforts to reduce inequalities, can be helped by careful description of the factors that might be important for transferring the evidence to specific populations. For example, populations may be characterised by their place of residence, religion, occupation, gender, Race/ethnicity, education, socioeconomic status, and social networks and capital (Evans and Brown 2003). This approach, with its mnemonic PROGRESS, for capturing social determinants of health, has been integrated into guidance for pre-specifying subgroup analyses in systematic reviews (Welch et al 2012; O'Neill et al 2014).

3.2 Synthesis Methods Innovation for International Development

In addition to choosing between existing synthesis methods, there is a need for a more inclusive research agenda that studies methods for evidence synthesis that, for example, includes methods that:

- aim to advance the methods of primary research, not just share the findings of rigorous primary research;
- are well developed in other academic disciplines but rarely applied to international development;
- take into account evidence about contextual influences of impact;
- expand the scope of evidence synthesis to include upstream research focused on developing effective interventions, not only evaluating them; [evaluability]
- include knowledge held by organisations and individuals who implement or are offered evidence-informed interventions;
- advance learning for themes, sectors or geographical areas where studies addressing impact are limited in number and rigour;
- build on what is known about sustainability with rigorous empirical research
- support asking better questions, not just developing more reliable answers [include conceptual frameworks, non-directive counselling, co-constructive learning]; and
- accrue learning from studies addressing how evidence is used.

3.2.1 Advancing the Methods for Primary Research

Some evaluation questions require methods for which there is no consensus in the international development research community as to the best available approach. Papers being prepared in parallel to this one are discussing: the strengths and weaknesses of major evaluation approaches (Attanasio et al, unpublished); and what approaches would be most appropriate for contexts in which current primary methods are not applicable (Hargreaves et al, unpublished).

Methodology reviews have a role in raising the profile of common methodological flaws and potential solutions identified by systematic reviews of the literature. Such reviews can be followed by consensus meetings to inform guidelines for evaluation methods taking a fit-for-purpose approach, acknowledging that some contexts do not allow the application of some existing methods. The aim of such consensus development is to: clarify how to choose between existing evaluation methods; develop new evaluation methods in the field as necessary; and raise the quality of primary evaluations. Additional challenges and solutions may be identified by inviting stakeholder critique of existing evidence, although this is not often done.

3.2.2 Adapting Synthesis Methods from Elsewhere for International Development

As the production of systematic reviews has spread across academic disciplines and policy sectors, the methodology has encountered new challenges. Nursing studies

prompted more attention to qualitative research. Public health was where mixed methods synthesis began with barriers and facilitators reviews. International development emphasised causal chain analysis. As the number of systematic reviews addressing international development grows, there is an opportunity to adopt and adapt methods from other disciplines. Being multidisciplinary itself, through spanning agriculture, economics, education, environmental science, health and more, international development provides an opportunity to discuss how these methods vary and what fundamental principles or operational concepts they share across academic disciplines. A useful starting point would be to review the guidance currently available for systematic reviews to consider its applicability to international development. An appropriate forum for such discussions is the Global Evidence Synthesis Initiative (GESI, http://www.gesiinitiative.com/about-gesi) (Langlois et al 2015). GESI was launched to enhance the capacity of Low- and Middle-Income Countries (LMICs) in synthesising evidence, and using synthesised evidence to support practice and policy across disciplines. GESI will achieve these aims through supporting Evidence Synthesis Centres based in LMICs.

Some synthesis approaches are yet to be applied in international development. Examples include structural equation modelling approaches to meta-analysis, which are currently applied in the educational, social, behavioural, and medical sciences (Cheung MWL 2015) and Qualitative Comparative Analysis, for which there are worked examples addressing complex interventions in health (Candy et al 2013; Brunton et al 2014; Kahwati et al 2016).

Qualitative synthesis and mixed methods syntheses, which draw on a broader range of evidence, are also stronger in some disciplines. For instance, tools which assess the confidence in findings from systematic reviews of effectiveness studies (GRADE[4]) and findings from systematic reviews of qualitative studies (GRADE-CERQual) have been developed for making recommendations in health and social care (Lewin 2015). The testing of these tools across other disciplines addressing international development is currently limited.

In health care, efforts to build an accumulative research literature have been enhanced by harmonising key outcomes and measures within fields of research through the development of core outcome sets (Clarke 2007). The COMET Initiative has attracted support across health care and advocates outcome sets that indicate the minimum to be measured and reported in all clinical trials, audits of practice or other forms of research for a specific condition. Core outcome sets 'do not imply that outcomes in a particular study should be restricted to those in the core outcome set. Rather, there is an expectation that the core outcomes will be collected and reported to allow the results of trials and other studies to be compared, contrasted and combined as appropriate; and that researchers will continue to collect and explore other outcomes as well' (http://www.comet-initiative.org/about/overview). The need for core outcomes sets in international development is illustrated by a systematic review of mental health and psychosocial support interventions for people affected by humanitarian

[4] http://www.gradeworkinggroup.org/

emergencies (Bangpan and Dickson 2017). To evaluate impact, a statistical meta-analysis was applied whenever two or more studies presented data for the same outcomes. The broad range of outcomes and their inconsistent application in trials allowed for statistical meta-analyses addressing some outcomes (post-traumatic stress disorder, depression, conduct problems, functional impairment, prosocial behaviours, psychological distress, anxiety, emotional problems, hope, social support and somatic complaints) but not others (coping, grief, suicide, guilt, stigmatisation and resilience). Although the principle of core outcome sets has been applied successfully in health care, the methods may need adapting for the multidisciplinary field of international development.

Box 11: Research recommendations for harmonising outcomes and measures for international development reviews

We recommend methodological research, informed by the COMET initiative, to develop core sets of outcomes and measures to enhance the accumulative nature of development impact and learning.

3.2.3 Accounting for Differences in the Context of Primary Studies

The reporting guidelines for addressing equity in systematic reviews (Welch et al 2012; O'Neill et al 2014) is readily applicable to public health because it provides a framework for epidemiological analyses. However, it is limited in terms of the range of potential contextual influences that it can capture. In terms of the umbrella model of social determinants of health, it disregards both proximal and distal factors that may vary internationally: the inner layers of individual risk factors (such as genetics, physical impairment or lifestyle factors) that feature in biology and behavioural science; and the outer layers of ecological or geological factors central to environmental science. Moreover, the PROGRESS tool takes no account of social identities overlapping or intersecting (Collins 2015), perhaps because multiplying subgroup analyses reduces statistical power in epidemiology (Burke et al 2015). Lastly, it ignores changes over time that may be particularly important for longitudinal studies: interactions over time between the multiple layers; the life course (age); life transitions (moving home, employment, school or leaving prison, hospital or a significant relationship); historical changes (conflicts, mass migrations, (post)colonialism); or geological or climate changes (natural disasters). An important example is the development of antibiotic resistance spreading over time which may limit the effectiveness of therapy inconsistently over time and geographical setting. This has been accommodated in a systematic review of treatment for malaria by documenting the year and country of each study (Sinclair et al 2009).

A more flexible approach to investigating contextual influences or inequalities may be found in the work of Bronfenbrenner (1979; 1995) who conceptualised children's lives as being shaped by environmental factors acting and interacting in a set of nested structures, from within families (at the micro level) to within their historical context (at the macro level). This has been applied to systematic reviews of research (Ang 2014)

and policy (Ang and Oliver 2015) addressing children's rights in post-conflict areas. The potential for applying such a framework generically to different systematic reviews is suggested by the various adaptations of ecological frameworks that can be found for primary research elsewhere, such as: environmental science (Coutts and Hahn 2015); migration studies (Nkulu Kalengayi et al 2012); and violence (Krug et al 2002).

Ecological models not only offer a framework to make sense of review findings but, as they provide a way to clarify the diversity of people's life circumstances, they also provide a framework for identifying stakeholders who can help with shaping the review or interpreting the findings. An ecological framework can be immensely beneficial when researching context-sensitive topic areas such as children, gender and the broader social, cultural environments.

However, the potential for ecological models to frame syntheses of qualitative research or inform the choice of sub-group analyses in syntheses of quantitative research has yet to be tested.

Box 12: Addressing contextual factors and inequalities in systematic reviews

We recommend systematic reviewers explore the utility of inequalities frameworks (e.g. the PROGRESS tool (Welch et al 2012; O'Neill et al 2014) and ecological models similar to Bronfenbrenner (1979; 1995) to take into account contextual factors in systematic reviews of qualitative and quantitative research.

3.2.4 Addressing Formative Evaluations, Not Just Summative Evaluations

Evidence synthesis has long brought together the findings of summative evaluations to assess the impact of interventions. More recently advances in synthesis that accommodates qualitative and mixed methods research designs have complemented this evidence of impact with findings to explain variations and disappointing impact (see Box 9 above for an example). These advances provide an opportunity to draw on the same studies not to explain impact, but as formative evidence in the design of interventions prior to their evaluation in rigorous impact studies. This requires knowledge about acceptable and feasible interventions, theories of change and numerical estimates of population characteristics in order to design, commission and implement impact evaluations in the field is commonly applied in the development of individual interventions, where it is considered important for assessing the readiness of an intervention for evaluation – its evaluability. 'Evaluability assessment involves key policymakers, managers, and staff in developing program theory and clarifying intended uses of evaluation information, thus helping solve problems that inhibit useful program evaluation' (Wholey 1987). This concept underpins guidance for evaluating complex interventions in health care (Craig et al 2014), and in international development (Davies 2013). There is an opportunity to translate this concept from primary research to research synthesis so that the development and evaluation of interventions can be informed by a wider body of knowledge. Such an expansion would embrace the evaluation criteria of development assistance first laid out by the OECD's Development

Assistance Committee (1991): relevance, effectiveness, efficiency, impact and sustainability.

In the past, primary research alone, such as document reviews, site visits, and interviews have constituted formative evaluations to determine the readiness of a specific programme for impact assessment, in other words, evaluability assessment (Trevisan 2007). The availability of synthesis methods that now accommodate a broader range of study designs presents opportunities to conduct formative evaluations and evaluability assessments by drawing on multiple studies addressing the development of goals and objectives, programme logic models, and modification of programme components. Although we are not aware of a systematic review explicitly synthesising formative assessments, discussions with the authors of a synthesis of qualitative research to develop theory allowed the evidence to be re-cast as a synthesis for developing interventions (See Box 13). This creates an opportunity for systematic reviews to inform adaptive management by drawing on emerging findings from multiple examples of evolving programmes.

Box 13: Dual purpose of synthesising qualitative research: theory development and intervention development (Example from Peters et al 2016).

Review questions: What are the perceived or apparent benefits/negative consequences of participating in a microfinance programme? What are people's motivations for deciding to participate/not participate, or drop out of a microfinance programme? What are the beliefs and attitudes of other household and community members towards microfinance-programme participants?

Learning for theory development: Many experiences of microfinance can be understood as constituted by three interrelated areas: empowerment; social and personal impact; and financial and economic impact. Household and community attitudes and beliefs contribute to how beneficiaries experience microfinance programme participation within these three areas. Ultimately, it appears that the interplay between beneficiaries' positive and negative experiences and the views of other household and community members is an important factor for how beneficiaries decide to initiate, continue or terminate their microfinance programme participation. Positive experiences of microfinance appear to be based on three fundamental principles:

Microfinance should enable the most vulnerable to become co-contributors to their family and community where empowerment is intrinsic to the positive experiences of women in particular.

Microfinance programmes should aim to include strategies that assist beneficiaries to develop social capital through the development of relevant skills and knowledge.

Financial and economic management experience is cumulative and has a key role in improving beneficiaries' self-confidence, their households' level of support, and positive wider community sentiment.

Without the integration of these themes, beneficiaries may be more vulnerable to negative experiences and outcomes such as loss of family/community support or social status.

Learning for intervention development: This synthesis and contextual analysis offers a number of implications for designing and implementing microfinance programmes:

Designing financial products to suit a range of circumstances; such as women earning their education; women working in agriculture (e.g. micro-insurance); and alternatives to microcredit where interest-bearing loans do not suit social norms.

Ensuing microfinance product repayment schedules are responsive to local environmental challenges and changes, particularly during times of famine, or disaster.

Encouraging group lending or self-help groups to offer opportunities for socialising and learning together to build a cohesive and supportive community.

Advertising the full range of potential benefits to attract a variety of clients.

Working with supportive husbands as role models for other men.

Addressing local hierarchies among beneficiaries and communities with good management, including formal decision-making processes, to ensure accountability to the community and equity amongst members.

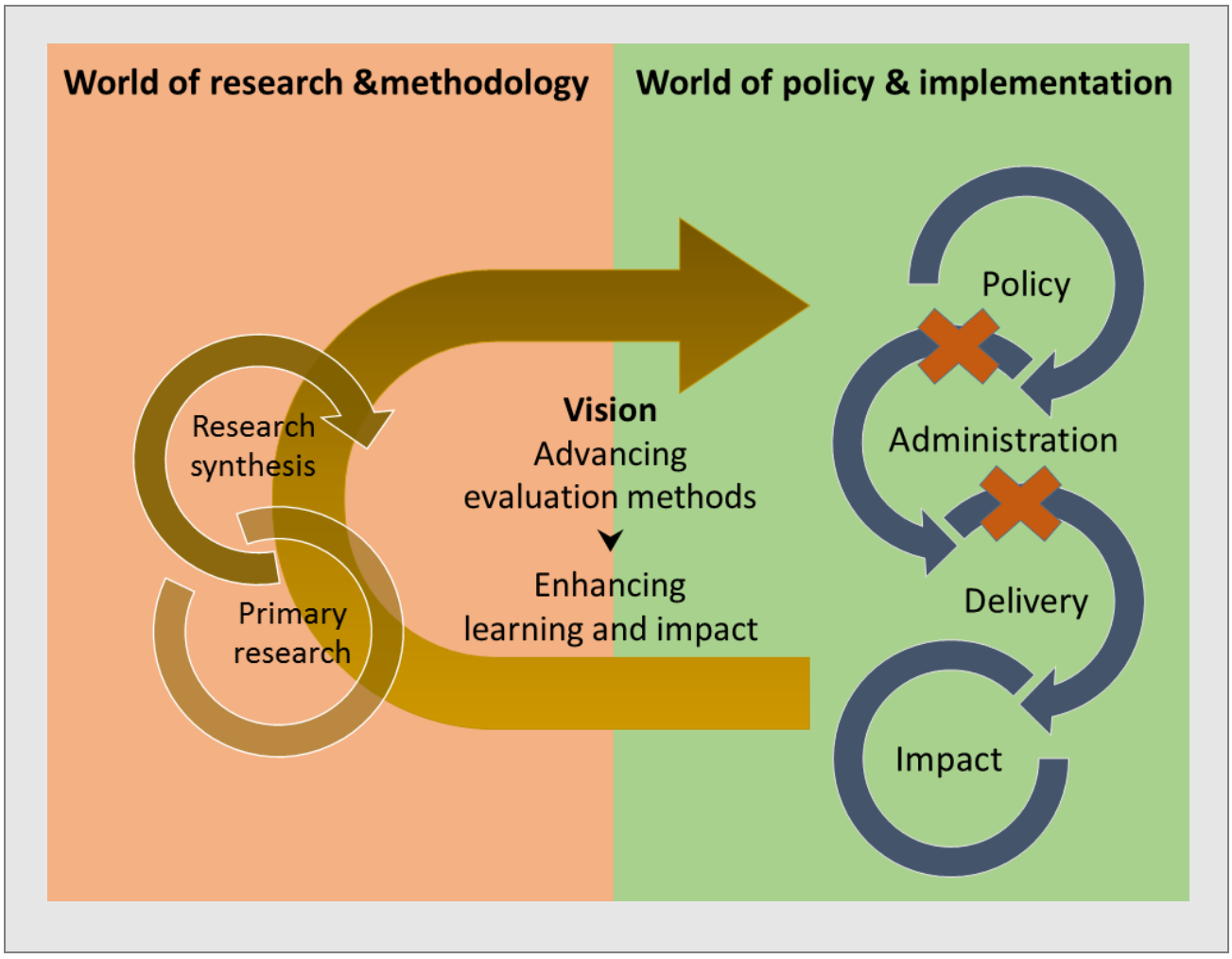
Box 14: Synthesising formative research

We recommend exploring the feasibility of synthesising formative research within on-going development programmes by integrating programme documents into methods for synthesising qualitative and mixed methods research with methods.

3.2.5 Synthesis Encompassing Organisational and Experiential Knowledge

Currently, evidence syntheses predominantly include academic research collated by teams of researchers with little or no input from people who might draw on syntheses for policy decisions, or administer or deliver programmes based on those decisions. Their content is largely evaluations of programmes, sometimes adapted by researchers specifically to enable rigorous evaluation and often stripped of their organisational and socio-political context. Consequently, the findings of these studies, with high internal validity, offer persuasive evidence of impact for policy decisions. Yet, the partial picture this evidence presents does little to inform decisions for implementing supposedly effective programmes at the level of policy administration or to provide contextual information to support programme delivery in the field. The result is evidence-informed policy decisions that subsequently stall and fail to deliver better outcomes. This situation is illustrated in figure 3 and discussed at greater length with examples from tuberculosis treatment programmes elsewhere (Oliver 2016).

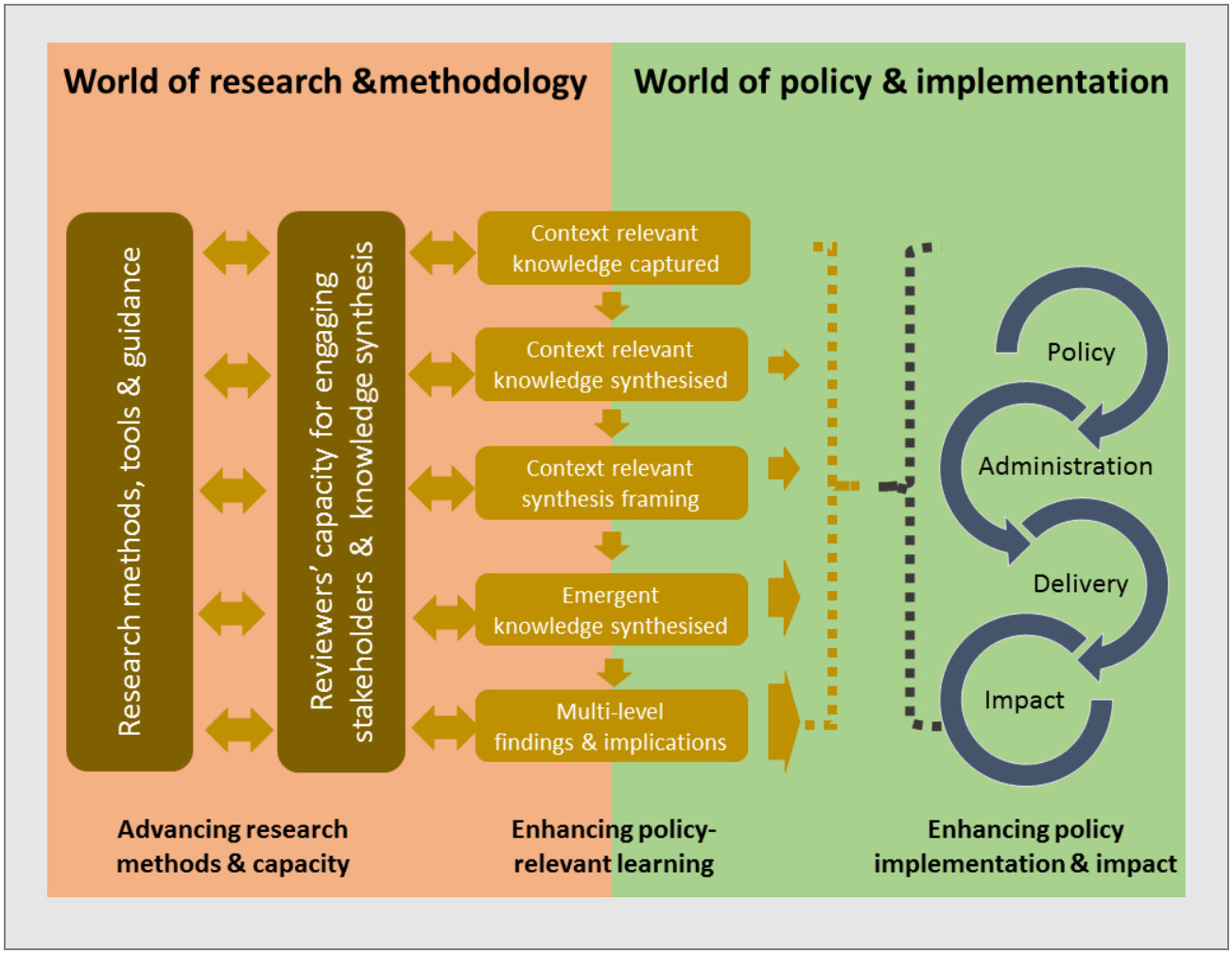
Figure 3: Typical limitations of knowledge transfer between worlds of policy and research (adapted from Oliver et al (submitted to Environmental Evidence) Research-based information about the impact of services flows from where it is collected (bottom right), typically from practice areas where data are framed by research tools and analysed to maximise the internal validity of primary studies (bottom left), and then synthesised to emphasise average effects with an assessment of the degree of heterogeneity of studies and judgements about generalisability of findings. Subsequently, summaries of syntheses are presented to panels, such as guideline groups, making policy decisions (top right). Information flow from policies to guide research base practice is interrupted during implementation efforts where evidence maximising external validity is required for systems issues, to complement evidence addressing practice issues (middle right).



Possible solutions come from implementation science which 'seeks to understand and work in 'real-world' or usual practice settings, paying particular attention to the audience that will use the research, the context in which implementation occurs, and the factors that influence implementation' (Peters et al 2013). Unlike implementation science, which focuses on promoting the adoption and integration of practices, interventions and policies underpinned by evidence, the aim here is to gather and integrate organisational and experiential knowledge earlier in order to inform the development of interventions and impact evaluations, and later when interpreting the findings of evaluations. Recent advances in evidence synthesis present an opportunity

to synthesise organisational and experiential knowledge from existing studies in different contexts to predict likely challenges and possible solutions to implementation (intervention or system adaptations) so that impact seen in a research context is replicated in the real world. This vision is illustrated in figure 4.

Figure 4: Potential advances in synthesis to enhance impact in the 'real world' (adapted from Oliver, Garner and Heywood (2015), unpublished discussions)



This vision recognises the value of involving various stakeholders from policy, administration or delivery organisations and populations where impact is sought (right-hand column of figure 4) in guiding systematic reviews. These stakeholders have the potential to influence review production (middle column of figure 4) by highlighting issues, or even studies, relevant to their interests or experiences and discussing how the synthesis may be framed and the findings interpreted. Responding constructively to this input requires review teams to have the capacity to facilitate such discussions and synthesise knowledge in ways that enable findings to be used taking into account the context of implementation. This relies on the capacity of individuals' skills, the capacity of teams to work together productively, the capacity of their host organisations to support their work, and the capacity of current research methods, tools and guidance to synthesise relevant knowledge (Oliver et al 2015).

Thus, this vision links the direct experiences of both professionals directing or delivering services and people offered or receiving services with state of the art methods for research synthesis. Advances currently required include: capturing knowledge relevant to various stakeholders from a real-world context, not only a research context; preparing syntheses that are sensitive to variation in contexts for using research findings; and syntheses of formative research and on-going studies to predict and adapt to contextual challenges arising from organisations delivering interventions and the locations where they are offered. In essence, the vision is to transform systematic reviews from the paradigm of science alone (mode 1), characterised by academically driven theoretical and experimental science to a new paradigm of knowledge production ('Mode 2'), which is socially distributed, application-oriented, trans-disciplinary, and subject to multiple accountabilities (Gibbons et al 1994; Nowotny et al 2001).

Gathering organisational and experiential knowledge, values and priorities
Documentary evidence about programmes painstakingly gathered through interviews and site visits is commonly available on organisations' websites. With 'web scraping' software now able to turn web pages into data a much broader sweep of evidence could be made available for synthesis to predict obstacles and solutions to implementation, and thereby support impact. However, such advances in IT have yet to be applied and evaluated in the context of evidence synthesis.

Synthesising on-going and completed studies: Extending synthesis to formative research opens opportunities for learning from pilot studies and emerging findings of projects across on-going programmes of work and more interactive ways of working between research and policy and practice development. CEDIL's location within DFID creates an exceptional opportunity to shape the process of accumulating evidence by working closely with DFID funded programmes. Two innovative strategies could be explored: interpreting the learning from interim reports from on-going projects in light of wider syntheses of evidence; and synthesising the learning from the same interim reports with the evidence accrued more widely. Both strategies aim to accelerate the shared learning and facilitate adaptive management within on-going programmes. Synthesising the findings of on-going studies, combined with rapid review methods, offers new opportunities for adaptive management. This, however, requires the development of methods for more detailed understanding of the analytical assumptions in using interim evidence to inform adaptive management (which relates to the issues of complexity and mixing methods discussed by Humphreys and Jacobs, 2015).

Box 15 Engaging stakeholders with responsibility for policy implementation

We recommend systematic reviewers exploring the feasibility of engaging stakeholders with responsibility for policy implementation in the tasks of framing systematic reviews and interpreting the findings.

3.2.6 Developing Methods to Address Thematic and Geographic Evidence Gaps

Section 3.1.4 introduced systematic maps as a method for navigating the literature and identifying substantive gaps in evidence. Some of these gaps (such as climate change, conflict and humanitarian settings, governance and infrastructure) are areas where extant studies are limited in number and more likely to be reflective case studies than controlled trials. Addressing these substantive gaps are likely to require methods for synthesising formative evaluations, organisational knowledge and on-going studies.

Box 16: Mapping the applicability of current synthesis methods to substantive areas

We recommend checking the methodological challenges for synthesis in substantive areas where synthesis has, as yet, had little application.

3.2.7 Sustainability

Although sustainability is widely seen as a key component of implementation models (Aarons et al 2011), formal assessment of sustainability is relatively rare. A systematic review of intervention sustainability in health care found studies rarely defined or rigorously evaluated sustainability (Wiltsey Stirman et al 2012). There is a role for synthesis in learning about sustainability models and evaluation methods. The examples mentioned here were developed in public health and may benefit from further development for wider application.

Box 17: Mapping the applicability of current synthesis methods to substantive areas

There is a need for exploring the applicability of sustainability frameworks to synthesis methods.

3.2.8 Asking Better Questions

Systematic reviews in international development have been driven more by policymakers than by academics, with policymakers asking questions which require a broader sweep of the literature. One solution has been two-stage reviews that map the available literature to inform a discussion which seeks to balance the most important sub-questions with sub-sections of the literature that are most likely to provide reliable evidence. Discussions for refining review questions, whether employing two-stage reviews or not, involve lateral thinking and constructive conflict to navigate the research-policy interface (Oliver et al, 2017). Generally, this initial stage in conducting reviews has received very little attention. Existing research has addressed framing questions with reviewers and a single policy team (Oliver et al 2017), sometimes with the support of knowledge brokers (Campbell et al 2011), for tailored reviews rather than public goods reviews. Involving broader groups of policymakers has been the focus of

deliberative debate once systematic reviews have been completed (Moat et al 2014), rather than when framing the questions. This work overlaps with what is known about setting research priorities (Viergever et al 2010; Akl, under peer review).

Box 18: Framing questions for public goods systematic reviews

There is a need to develop methods that involve collaborative groups of policymakers and evidence providers to discuss theories of change and how to generate better questions from them.

3.2.9 Accruing Learning from Studies Addressing How Evidence is Used

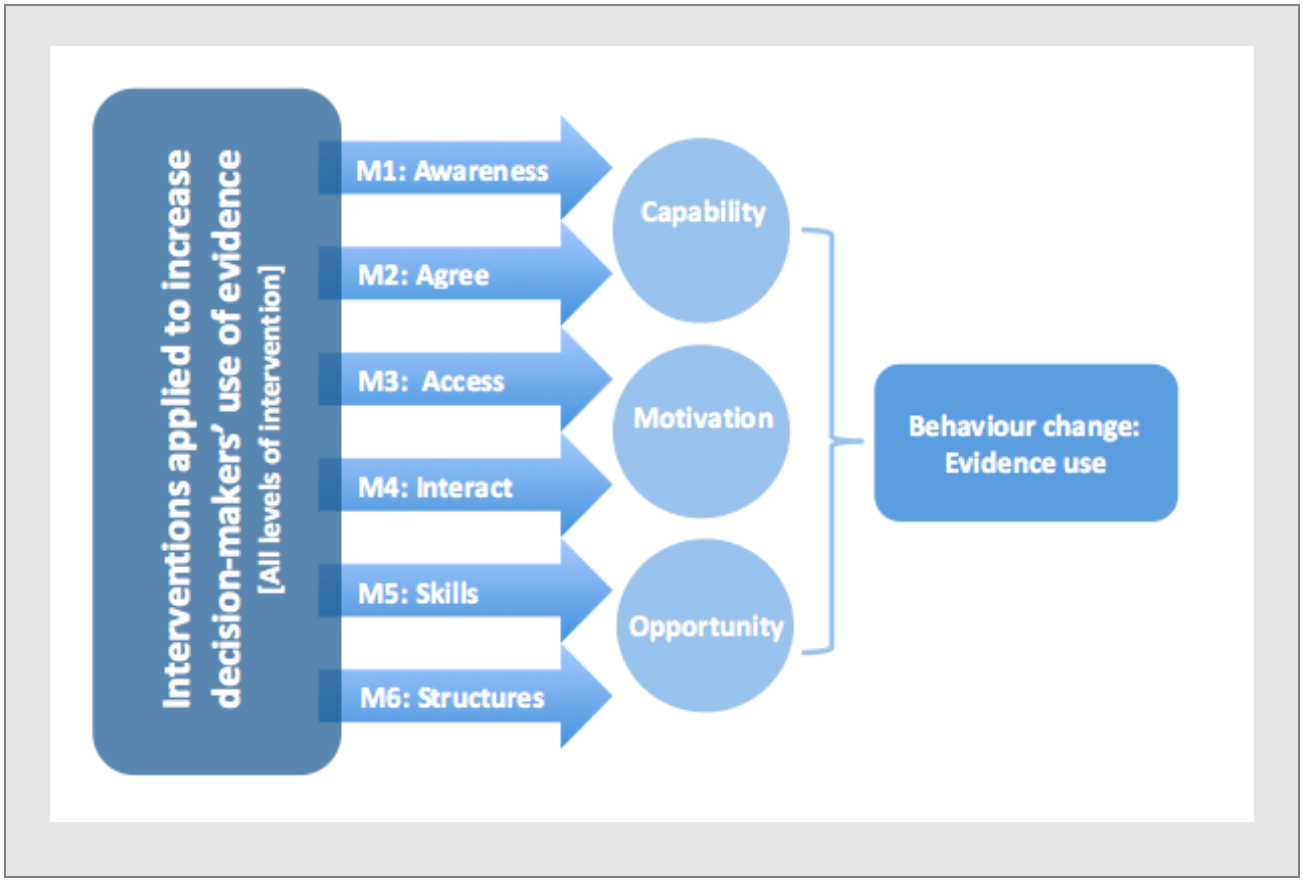
This paper has argued for the use of formal explicit methods for clarifying both what we know to inform policy making and also what further primary research is required to inform future decision making. The paper has also discussed some of the complications of the evidence being fit for purpose and relevant to decision makers in international development. This raises the issue of how evidence is considered and used which is itself an increasing area of research, particularly systematic reviews. This shows that the assumption that research is available and therefore should be used is often not the case. Barriers to research use commonly include poor access to good quality relevant research and lack of timely research output; and the facilitators most often reported are collaboration between researchers and policymakers, and improved relationships and skills (Kathryn Oliver et al 2014). People outside of academia may also want to (from Chapter 11 in Gough et al 2017):

- Be aware of the research
- Have access to published work
- Be able to understand the findings
- Be able to check the trustworthiness of the findings
- Be able to check how the findings fit into what else is known about the issue
- Be able to assess the relevance of these and other findings to their own contexts

As well as research on understanding how and why research is or is not used, there are also a range of strategies to assist research use through the development of policy and practice guidance based on formal deliberative procedures that considers the research, the various perspectives concerned with the decisions being made and non-research experiential or practice knowledge. A recent systematic review summarised the features of interventions that were shown to successfully increase the use of evidence (Langer et al 2016). The review used a framework that allowed intervention components to be understood in terms of the level of intervention, single and multi-mechanisms and components of behaviour change of capability, motivation, and/or opportunity to use evidence (Figure 5). The review found that there was more evidence of effective interventions when behavioural components such as strategies to increase motivation to use evidence were added to the use of single mechanisms of change (such as the

often-recommended strategy of developing interventions between researchers and users of evidence (Langer et al 2016).

Figure 5: Intervention logic model for use of evidence (Langer et al 2016)



We need methods to map what research has been undertaken, methods to synthesise the evidence to inform decision-makers, methods to incorporate synthesised research with other evidence and perspectives, and methods to synthesise the evidence on how research is used. This last issue is attracting growing interest (see for example, Gough and Tripney 2016). For instance, although a framework is available to support decision-making based on systematic reviews for healthcare (Alonso-Coello et al 2016) such work would benefit from further attention, taking into account the full range of systematic reviews in the social sciences and the degree to which they can accommodate the ambiguity and complexity of the individual studies they examine.

Section 4

Research Agenda

Based on the discussions above we propose a research agenda to include the following:

An open call for methodological research

As innovations so far have built on the experience of teams addressing substantive research questions, an open research call for methods development should attract proposals from experienced teams. This is now included in the research agenda.

Specific priorities for research include the following.

Distinguishing public goods reviews from tailored reviews, where

- Research funders invest in producing and cataloguing 'public goods' systematic reviews in an open-access global evidence repository (Model 1).
- Decision-making organisations raise awareness and skills for their staff to draw on such a global repository of systematic reviews (Model 1).
- When synthesised evidence is required for urgent decisions, systematic methods be applied (Models 2 and 4), and systematic reviewers draw on these repositories to use or re-analyse existing systematic reviews containing relevant evidence.
- Investigating the relationship between public goods reviews and tailored reviews.

When evidence is synthesised fast (models 2 and 4) owners of systematic review repositories investigate opportunities to transform this evidence into 'public goods' systematic reviews (models 1 and 3).

Developing methods guidance

Guidance for research synthesis, whether aimed at producers or users of reviews should clarify:

- the essential principles that underpin all approaches to research synthesis
- how approaches and detailed methods can vary for different purposes

how to choose appropriate methods depending on the question and literature available

Transdisciplinary research methods

Systematic reviewers and their review partners should draw together their collective experience of working across and beyond academic boundaries on an international scale and develop guidance for their peers.

Evidence gap maps

CEDIL, in discussion with curators of evidence libraries and evidence gap maps, should consider the options available for navigating evidence sources by different potential users. This should be done in discussion with potential users whether they have responsibility for specific programmes, specific populations or specific geographical areas.

Harmonising outcomes and measures for international development reviews

We recommend methodological research, informed by the COMET initiative, to develop core sets of outcomes and measures to enhance the accumulative nature of development impact and learning.

Addressing contextual factors and inequalities in systematic reviews

We recommend systematic reviewers explore the utility of inequalities frameworks (e.g. the PROGRESS tool (Welch et al 2012; O'Neill et al 2014) and ecological models similar to Bronfenbrenner (1979; 1995) to take into account contextual factors in systematic reviews of qualitative and quantitative research.

Synthesising formative research

We recommend exploring the feasibility of synthesising formative research within ongoing development programmes by integrating programme documents into methods for synthesising qualitative and mixed methods research with methods.

Engaging stakeholders with responsibility for policy implementation

We recommend systematic reviewers exploring the feasibility of engaging stakeholders with responsibility for policy implementation in the tasks of framing systematic reviews and interpreting the findings.

Mapping the applicability of current synthesis methods to substantive areas

We recommend checking the methodological challenges for synthesis in substantive areas where synthesis has, as yet, had little application.

Mapping the applicability of current synthesis methods to substantive areas

There is need to explore the applicability of sustainability frameworks to synthesis methods.

Framing questions for public goods systematic reviews

There is a need to develop methods that involve collaborative groups of policymakers and evidence providers to discuss theories of change and how to generate better questions from them.

Acknowledgements

We are grateful to Nancy Cartwright, Rick Davies, Andrew Pullin and Howard White for their comments on earlier versions of this paper. We also drew on discussions within the CEDIL Intellectual Leadership Team.

Funding

We are grateful for funding from the Department for International Development. The views expressed do not necessarily reflect the UK Government's official policies.

References

Aarons GA, Hurlburt M, McCue Horwitz S (2011) Advancing a Conceptual Model of Evidence-Based Practice Implementation in Public Service Sectors. Administration and Policy in Mental Health and Mental Health Services Research 38 (1) 4 – 23.

Akl EA, Fadlalla2 R, Ghandour L, Kdouh O, Langlois E, Lavis JN, Schünemann Holger, El-Jardali F (under peer review) The SPARK Tool for Prioritizing Questions for Systematic Reviews in Health Policy and Systems Research: Development and initial Validation.

Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, Treweek S, Mustafa RA, Rada G, Rosenbaum S, Morelli A, Guyatt GH, Oxman, AD, the GRADE Working Group (2016) GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *British Medical Journal* 353:i2016.

Anderson, E., Jalles d'Orey M, A., Duvendack, M., Esposito, L. (2016) What policies and interventions have been strongly associated with changes in in-country income inequality? London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London Available at: https://assets.publishing.service.gov.uk/media/57bb113aed915d1257000002/Systematic_Review_CHANGES_IN_INCOME_INEQUALITY.pdf

Anderson E, Jalles d'Orey MA, Duvendack M, Esposito L (2016) What policies and interventions have been strongly associated with the translation of growth into reductions in income poverty? London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London available at: https://assets.publishing.service.gov.uk/media/57bb1efae5274a096b00000c/TRANSLATION_OF_GROWTH_INTO_REDUCTIONS_IN_INCOME_POVERTY.pdf

Ang L. *Early Childhood and Peace Building in the Asia-Pacific Region: A Literature Review to Inform a Regional Research Agenda*, UNICEF Publication, 2014. Accessed 1 April 2017 https://www.arnec.net/wp-content/uploads/2015/04/Peace-Bldg-Report-ref.pdf

Ang L, Oliver S. *A Systematic Policy Review of Early Childhood Development and Peacebuilding in fourteen Conflict-affected and Post-conflict countries.* UNICEF and UCL Institute of Education: University College London, 2015. Accessed 1 April 2017

Annamalai TR, Devkar G, Mahalingam A, Benjamin S, Rajan SC, Deep A (2016) What is the evidence on top-down and bottom-up approaches in improving access to water, sanitation and electricity services in low-income or informal settlements? London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

Attanasio O, Rasul I, Cavatorta E (unpublished) *Major evaluation approaches and their strengths and weaknesses*. CEDIL, London.

Bangpan M, Dickson K, Felix L, Chiumento A. (2017) *The impact of mental health and psychosocial support interventions on people affected by humanitarian emergencies: A systematic review*. Humanitarian Evidence Programme. Oxford: Oxfam GB.

Barakat S, Hardman F, Connolly D, Sundaram V, Zyck SA. (2010) *Programme review & evaluability study (PRES) UNICEF's education in emergencies & post-crisis transition (EEPCT) programme*. UNICEF and Postwar Reconstruction and Development Unit, University of York.

Berendes S, Heywood P, Oliver S, Garner P. Quality of private and public ambulatory health care in low and middle income countries: systematic review of comparative studies. *PLoS Med*. 2011 Apr;8(4):e1000433. Epub 2011 Apr 12.

Booth A and Carroll C (2015) How to build up the actionable knowledge base: The role of 'best fit' framework synthesis for studies of improvement in healthcare. *BMJ Quality & Safety*, 24(11): 700–708.

Boulton, J (2015) *Embracing complexity: strategic perspectives for an age of turbulence*. Oxford: Oxford University Press.

Brunton G, O'Mara-Eves A, Thomas J. (2014) The 'active ingredients' for successful community engagement with disadvantaged expectant and new mothers: a qualitative comparative analysis. J Adv Nurs. 70(12):2847–60. doi:10.1111/jan.12441.

Bronfenbrenner U. *The Ecology of Human Development*. Cambridge, MA: Harvard University Press, 1979.

Bronfenbrenner U. 'Developmental Ecology through Space and Time. A Future Perspective' in P. Moen & G. H. Elder, Jr., (Eds.), *Examining lives in context: Perspectives on the ecology of human development* (pp. 619-647). Washington, DC: American Psychological Association, 1995.

Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses *BMJ* 2015; 351: h5651.

Campbell D, Donald B, Moore G, Frew D. (2011) Evidence check: knowledge brokering to commission research reviews for policy. *Evidence and Policy.* 7:97–107.

Candy B, King M, Jones L and Oliver S (2013) Using qualitative evidence on patients' views to help understand variation in effectiveness of complex interventions: a Qualitative Comparative Analysis. *Trials* **14**:179 doi:10.1186/1745-6215-14-179 http://www.trialsjournal.com/content/14/1/179

Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM, Howells DW, Ioannidis JPA, Oliver S. (2014) How to increase value and reduce waste when research priorities are set. *The Lancet* 382: 156 – 165.

Cheung MWL (2015) *Meta-analysis: a structural equation modeling approach.* Wiley.

Choi and Pak (2006) Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness*. Clin Invest Med* 29 (6) 351 – 364.

Choi and Pak (2007) Multidisciplinarity, interdisciplinarity, and transdisciplinarity in health research, services, education and policy: 2. Promotors, barriers, and strategies of enhancement. *Clin Invest Med.* 30 (6): E224-E232.

Clarke M (2007) Standardising outcomes for clinical trials and systematic reviews. *Trials* 8: 39.

Collins PH. Intersectionality's Definitional Dilemmas. *Annual Review of Sociology* 2015; 41:1–20.

Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Pettricrew M (2008) Developing and evaluating complex interventions: The new Medical Research Council guidance. *British Medical Journal 337 doi: https://doi.org/10.1136/bmj.a1655 (P* Coutts C, Hahn M. Green Infrastructure, Ecosystem Services, and Human *Health Int. J. Environ. Res. Public Health* 2015; 12(8), 9768-9798; doi:10.3390/ijerph120809768

Davies R (2013) *Planning evaluability assessments: a synthesis of the literature with recommendations.* Working paper 40, DFID.

Development Assistance Committee (1991) The DAC Principles for the Evaluation of Development Assistance, OECD.

DFID (2016) DFID Research Review, London.

DFID (2016) *Terms of Reference: Research Director and Intellectual Leadership - Centre of Excellence for Development Impact and Learning*. PO 7543 – ITT Volume 3.

Dixon-Woods M (2011) Using framework-based synthesis for conducting reviews of qualitative studies. *BMC Med*. 2011 Apr 14;9:39. doi: 10.1186/1741-7015-9-39.

Duvendack M, Hombrados JG, Palmer-Jones R, Waddington H. (2012) Assessing 'what works' in international development: meta-analysis for sophisticated dummies, *Journal of Development Effectiveness*, 4:3, 456-471, DOI: 10.1080/19439342.2012.710642.

Eddy-Spicer D, Ehren M, Bangpan M, Khatwa M, Perrone F (2016) Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised? A realist synthesis of school accountability in low-and middle-income countries. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London. Available at:
https://assets.publishing.service.gov.uk/media/5851931140f0b60e4c0000bd/SR_Q4_Final_Draft_for_Publication.pdf

Evans T, Brown H. Road traffic crashes: operationalizing equity in the context of health sector reform. Injury Control Safety Promotion **2003; 10**(1-2), 11-12.

Evans T, Brown H. Road traffic crashes: operationalizing equity in the context of health sector reform. Injury Control Safety Promotion 2003; 10(1-2), 11-12.evqanLavis, J, 2009, How Can We Support the Use of Systematic Reviews in Policymaking? Plos Medicine Nov; 6(11): e1000141.

Featherstone RM, Dryden DM, Foisy M, Guise JM, Mitchell MD, Paynter RA, Robinson KA, Umscheid CA, Hartling L. (2015) Advancing knowledge of rapid reviews: an analysis of

results, conclusions and recommendations from published review articles examining rapid reviews. *Systematic Reviews* 4:50. DOI: 10.1186/s13643-015-0040-4

Fischer K, Ekener-Petersen E, Rydhmer L and Edvardsson Björnberg K (2015) Social Impacts of GM Crops in Agriculture: A Systematic Literature Review. *Sustainability* 2015, 7, 8598-8620; doi:10.3390/su7078598

Flyvbjerg, B (2001) *Making social science matter: why social inquiry fails and how it can succeed again.* Cambridge: Cambridge University Press.

Garner P, Hopewell S, Chandler J, MacLehose H, Akl E, Beyene J, Chang S, Churchill R, Dearness K, Guyatt G, Lefebvre C, Liles B, Marshall R, Martínez García L, Mavergames C, Nasser M, Qaseem A, Sampson M, Soares-Weiser K, Takwoingi Y, Thabane L, Trivella M, Tugwell P, Welsh E, Wilson E, Schünemann HJ (2016) When and how to update systematic reviews: consensus and checklist. *British Medical Journal* 354 i3507.

G20 Germany (2017) Priorities of the 2017 G20 summit. Hamburg. https://www.g20.org/Content/DE/_Anlagen/G7_G20/2016-g20-praesidentschaftspapier-en.pdf?__blob=publicationFile&v Accessed 17 April 2017.

Gibbons M, Limoges C, Nowotny H, Schwartzman S, Scott P, and Trow M. (1994) The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies (London: Sage)

Gough D. Oliver S, Thomas J (2017) *Introduction to Systematic Reviews*, Second Edition. London: Sage.

Gough D, Tripney JS. (2016). Systematic reviews for policy. In G. Stoker, M. Evans (Eds.), *Evidence-based policy making in the social sciences: methods that matter* (pp. 43-67). Bristol: Policy Press.

Greenhalgh T, Robert G, Macfarlane F, et al. (2005) Storylines of research in diffusion of innovation: A meta-narrative approach to systematic review. Soc Sci Med 61: 417-430.

Grint, K (2005) Problems, problems, problems: the social construction of 'leadership'. *Human Relations*, 58(11):1467-1494.

Harden A, Thomas J (2010) Mixed methods and systematic reviews: examples and emerging issues. In '*The SAGE handbook of mixed methods in social and behavioural research*' edited by A Tasheakkori and C Teddlie. Chapter 29, pages 749-774. SAGE: Los Angeles (second edition)

Hargreaves J, Bonell C, Cartwright N, Prost A, Humphries M, Davey C, Hassan S (unpublished) *Contexts in which current methods are not applicable, discussing what approaches would be most appropriate.* CEDIL, London.

Higgins, S. and Katsipataki, M. and Villanueva-Aguilera, A.B. and Coleman, R. and Henderson, P. and Major, L.E. and Coe, R. and Mason, D. (2016) 'The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit', Manual. Education Endowment Foundation, London.

Higgins JPT, Lasserson T, Chandler J, Tovey D, Churchill R. (2016) Methodological Expectations of Cochrane Intervention Reviews. Cochrane: London.

Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. *Implementation Science*, 2010, 5:56

HM Treasury and DFID (2015) *UK aid: tackling global challenges in the national interest*. London.

Humphreys M, Jacobs AM (2015) Mixing methods: a Bayesian Approach. *American Political Science Review.* 109 (4) pp. 653-673

Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, Tibshirani R (2014) Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet* 383 (9912) 166–175.

The Joanna Briggs Institute (2014) *The Joanna Briggs Institute Reviewers' Manual 2014.*

Kahwati L, Jacobs S, Kane H, Lewis M, Viswanathan M Golin CE (2016) Using qualitative comparative analysis in a systematic review of a complex intervention. *Systematic Reviews* 5:82 DOI: 10.1186/s13643-016-0256-y

Koehlmoos T, Gazi R, Hossain S, Rashid M (2011) Social franchising evaluations: a scoping review. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Klumper W, Qaim M (2014) A Meta-Analysis of the Impacts of Genetically Modified Crops. *PLoS ONE* 9(11): e111629. doi:10.1371/journal.pone.0111629

Kneale D, Thomas J, Harris K (2015) Developing and Optimising the Use of Logic Models in Systematic Reviews: Exploring Practice and Good Practice in the Use of Programme Theory in Reviews. *PLoS ONE* 10(11): e0142187. https://doi.org/10.1371/journal.pone.0142187.

Kramer MS, Kakuma R. The optimal duration of exclusive breastfeeding: a systematic review. World Health Organization, 2002.

Krug EG, Dahlberg LL, Mercy JA, Zwi AB, Lozano R. (Eds) *World Report on Violence and Health* Geneva: World Health Organisation, 2002.

Langer L, Stewart R. (2013) What have we learned from the application of systematic review methodology in international development? – A thematic overview. *Journal of Development Effectiveness,* 6:3, 236-248, DOI: 10.1080/19439342.2014.919013.

Langer L, Tripney J, Gough D (2016) *The Science of Using Science: Researching the Use of Research Evidence in Decision-making*. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

Langlois E, Ranson K, Ghaffar A, Bärnighausen T, Bosch-Capblanch X, Daniels K, El-Jardali F, Grimshaw J, Haines A, Lavis J, Lewin S, Meng Q, Oliver S, Pantoja T, Straus S, Shemilt I, Tovey D, Tugwell P, Waddington H, Wilson M, Yuan B, Røttingen JA (2015) Advancing the field of health systems research synthesis. *Systematic Reviews* 4:90.

Lewin S, Glenton C, Munthe-Kaas H, Carlsen B, Colvin CJ, Gülmezoglu M, Noyes J, Booth A, Garside R, Rashidian A. (2015) Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from

qualitative evidence syntheses (GRADE-CERQual). *PLoS Med* 12 (10) 001895. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001895

Lewin S, Munabi-Babigumira S, Glenton C, Daniels K, Bosch-Capblanch X, van Wyk BE, Odgaard-Jensen J, Johansen M, Aja GN, Zwarenstein M, Scheel IB. Lay health workers in primary and community health care for maternal and child health and the management of infectious diseases. *Cochrane Database of Systematic Reviews* 2010, Issue 3. Art. No.: CD004015. DOI: 10.1002/14651858.CD004015.pub3.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. PLoS Med e6

Lopes Cardozo M et al. *Literature Review: Youth Agency, Peacebuilding and Education.* Amsterdam, Netherlands, Research Consortium Education and Peacebuilding, University of Amsterdam, 2015 (https://educationanddevelopment.files.wordpress.com/2014/11/youth-literature-review.pdf, accessed 22 May 2017).

Liu H, Muhunthan J, Hayek A, Hackett M, Laba TL, Peiris D, Jan S (2016) Examining the use of process evaluations of randomised controlled trials of complex interventions addressing chronic disease in primary health care—a systematic review protocol. Systematic Reviews 5: 138.

Masset E, Snilstveit B, Phillips D (in preparation) *A map of systematic evidence maps relating to lower-and middle income countries.*

Manski, C (2013) Predicting policy outcomes. In: Public policy in an uncertain world. *Cambridge, MA: Harvard University Press.*

Moat KA, Lavis JN, Clancy SJ, El-Jardali F, Pantoja T and for the Knowledge Translation Platform Evaluation study team (2014) Evidence briefs and deliberative dialogues: perceptions and intentions to act on what was learnt. *Bulletin of the World Health Organization* 92(1): 20–28.

Moe-Byrne T, Chambers D, Harden M, McDaid C (2014) Behaviour change interventions to promote prescribing of generic drugs: a rapid evidence synthesis and systematic review. *BMJ Open*, 4:e004623.

Moore G, Redman S, D'Este C, Makkar S, Turner T. (2017) Does knowledge brokering improve the quality of rapid review proposals? A before and after study. *Systematic Reviews* 6:23.

Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A and Volmink J (2007) Patient adherence to tuberculosis treatment: A systematic review of qualitative research. *PLoS Med*, 4(7): e238.

Nasser M, Welch V, Ueffi ng E, Crowe S, Oliver S, Carlo R. (2013) Evidence in agenda setting: new directions for the Cochrane Collaboration. *J Clin Epidemiol* 66: 469–71.

Natsios, A (2010). *The clash of counter-bureaucracy and development* Center for Global Development Essay. Washington DC.

Nkulu Kalengayi FK, Hurtig AK, Ahlm C, Ahlberg BM. "It is a challenge to do it the right way": an interpretive description of caregivers' experiences in caring for migrant patients in Northern Sweden. *BMC Health Services Research* 2012; 12 (433) DOI: 10.1186/1472-6963-12-433.

Nowotny, Helga; Peter Scott; Michael Gibbons (2001) Rethinking science: knowledge in an age of uncertainty. Cambridge: Polity. ISBN 0-7456-2607-6.

Obuku EA, Stewart R, Mijumbi R, Ochan M, Achana F, Akena D, Nakitende A, Ssemata A, Kinengyere A, Semakula D, Ssenono R, Nsangi A, Lalitha R, Mwesiga E, Akite J, Basaza R, Newbrander W, Okello J, Sewankambo N, with Dickson K, Oliver S. (2017) *Working with non-state providers in post-conflict and fragile states in primary healthcare service delivery: a systematic review.* EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

Oliver K, Innvaer S, Lorenc T, Woodman J, Thomas J (2014) *A systematic review of barriers to and facilitators of the use of evidence by policymakers.* BMC Health Services Research, 14: 2.

Oliver S*. Randomise until it hurts. International Journal* of *Epidemiology*, 2016, 321–326.

Oliver S, Bangpan M, Stansfield, Stewart R (2015) Capacity for conducting systematic reviews in Low and Middle Income Countries: a rapid appraisal. *Health Research Policy and Systems* 13**:**23 doi: 10.1186/s12961-015-0012-0

Oliver S, Dickson K (2016) Policy-relevant systematic reviews to strengthen health systems: models and mechanisms to support their production. *Evidence and Policy.* 12 (2) 235-259.

Oliver SR, Rees RW, Clarke-Jones L, Milne R, Oakley AR, Gabbay J, Stein K, Buchanan P, Gyte G. A multidimensional conceptual framework for analysing public involvement in health services research. *Health Expectations 2008*; 11 (1): 72–84

Oliver S, Wilson M, Melendez-Torres GJ, Bangpan M, Dickson K, Vigurs C (in press) Selecting rapid review methods for complex questions related to health policy and system improvements. In: Tricco A, Straus S, Langlois E (eds) *Guide to Rapid Reviews of Health Policy & Systems Research*. World Health Organisation.

Oliver S, Bangpan M, Dickson K (under peer review) Producing policy relevant systematic reviews: navigating the policy-research interface.

O'Neill (Petkovic) J, Tabish H, Welch V, Petticrew M, Pottie K, & Clarke M. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. Journal of Clinical Epidemiology 2014; 67, 56-64.

Pawson, R (2013) The science of evaluation: a realist manifesto.

Peters DH, Adam T, Alonge O, Akua Agyepong I, Tran N (2013) Implementation research: what it is and how to do it. *BMJ* 2013;347: f6753 doi: 10.1136/bmj.f6753.

Peters MDJ, Lockwood C, Munn Z, Moola S, Mishra RK (2016) People's views and experiences of participating in microfinance interventions: A systematic review of

qualitative evidence. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London. ISBN: 978-1-907345-87-6.

Petticrew M, Anderson L, Elder R, Grimshaw J, Hopkins D, Hahn R, Krause L, Kristjansson E, Mercer S, Sipe T, Tugwell P, Ueffing E, Waters E, Welch V (2015) Complex interventions and their implications for systematic reviews: A pragmatic approach. *International Journal of Nursing Studies* 52 (2015) 1211–1216.

Phillips, D, Coffey, C, Tsoli, S., Stevenson, J., Waddington, H., Eyers, J., Snilstveit, B. 2017. *A map of evidence maps relating to lower-and middle income countries*. 3ie, London.

*Pritchett, L and Sandefur J (*2015) Learning from Experiments When Context Matters*. American Economic Review, 105(5):471-75.*

Puzzolo E, Stanistreet D, Pope D, Bruce N, Rehfuess E (2013) Factors *influencing the large-scale uptake by households of cleaner and more efficient household energy technologies*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

The Sheffield Institute for International Development (2015) *A Hundred Key Questions for the Post-2015 Development Agenda.* United Nations Research Institute for Social Development and The Sheffield Institute for International Development. 4 http://id100.group.shef.ac.uk/include/A-Hundred-Key-Questions-for-the-Post-2015-Development-Agenda.pdf Accessed 17 April 2017.

Rose, P.; Battock, M. *Review of the DFID Systematic Review Programme*. DFID, London, UK (2012) 34 pp. https://www.gov.uk/dfid-research-outputs/review-of-the-dfid-systematic-review-programme Accessed 20 April 2017.

Sainfort F, Kuntz KM, Gregory S, Butler M, Taylor BC, Kulasingam S, Kane RL (2013) Adding Decision Models to Systematic Reviews: Informing a Framework for Deciding When and How to Do So. *Value in Health* 16 (1) 133-139.

Shaukat M et al., eds. *Screening newborns for congenital hypothyroidism (CHT)*. London, UK, Social Science Research Unit, UCL Institute of Education, University College London, 2015.

Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. (2007) Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 7:10. PMID: 17302989.

Sinclair D, Zani B, Donegan S, Olliaro P, Garner P. Artemisinin-based combination therapy for treating uncomplicated malaria. Cochrane Database of Systematic Reviews 2009, Issue 3. Art. No.: CD007483. DOI: 10.1002/14651858.CD007483.pub2.

Snilstveit, B, 2012, Systematic reviews: from 'bare bones' reviews to policy relevance. *Journal of Development Effectiveness* 4 (3) 388-408.

Steel Z, Marnane C, Iranpour C, Chey , Jackson JW, Patel V, Silove D (2014)The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *Int J Epidemiol*. 43(2): 476–493.

Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. JAMA; 283:2008–12.

Taylor S, Pinnock H, Epiphanou E, Pearce G, Parke H, Schwappach A, Purushotham N, Jacob S, Griffiths C, Greenhalgh T, Sheikh A.. A rapid synthesis of the evidence on interventions supporting self-management for people with long-term conditions: PRISMS - Practical systematic Review of Self-Management Support for long-term conditions. Health Services and Delivery Research 2014; 2(53).

Thomas J, Harden A, Oakley A, Oliver S**,** Sutcliffe K, Rees R, Brunton G, Kavanagh J. Integrating qualitative research with trials in systematic reviews: an example from public health. *British Medical Journal* 2004; 328: 1010-1012.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45.

Thomas J, Newman M, Oliver S. (2013) Rapid evidence assessments of research to inform social policy: taking stock and moving forward. *Evidence & Policy* 9 (1) 5–27.

Toye F, Seers K, Allcock N, Briggs M, Carr E, Barker K. (2014) Meta-ethnography 25 years on: challenges and insights for synthesising a large number of qualitative studies. *BMC Medical Research Methodology*201414:80. DOI: 10.1186/1471-2288-14-80

Trevisan MS (2007) Evaluability Assessment From 1986 to 2006. *American Journal of Evaluation* 28 (3): pp 290–303.

Tricco A, Straus S, Langlois E (eds) (in press) *Guide to Rapid Reviews of Health Policy & Systems Research*. World Health Organisation.

Vaessen, J (2016) Complexity in review and synthesis studies. In '*Dealing with complexity in development evaluation: a practical approach*', edited by M Bamberger, J Vaessen and E Raimondo. Ch.9, pages 165-180. SAGE: Los Angeles.

Viergever RF, Olifson S, Ghaffar A, Terry RF (2010) A checklist for health research priority setting: nine common themes of good practice. *Health Res Policy Syst* 8:3

Volmink J, Garner P (2006) Directly observed therapy for treating tuberculosis. *Cochrane Database of Systematic Reviews* 2. CD003343. doi:10.1002/14651858.CD003343.pub2. CD003343..

Waddington H, White H, Snilstveit B, Hombrados JG, Vojtkova M, Davies P, Bhavsar A, Eyers J, Perez Koehlmoos T, Petticrew M, Valentine JC, Tugwell P (2012) How to do a good systematic review of effects in international development: a tool kit, Journal of Development Effectiveness, 4:3, 359-387, DOI: 10.1080/19439342.2012.711765

Wang et al (2005) 'Applicability and transferability of interventions in evidence-based public health', *Health Promotion International* 21 (1): 76-83 2006

Welch V, Petticrew M, Tugwell P, Moher D, O'Neill J, Waters E, et al. PRISMA-Equity 2012 Extension: Reporting Guidelines for Systematic Reviews with a Focus on Health Equity. *PLoS Med 2012;* 9(10): e1001333. doi:10.1371/journal.pmed.100133.

Whiting P, Davies P, Savović J, CaldwellD, Churchill R. (2016) ROBIS: A new tool to assess risk of bias in systematic reviews was developed. Journal of Clinical Epidemiology 69: 225–234

WHO (2012) Optimizing health worker roles to improve access to key maternal and newborn health interventions through task shifting. World Health Organization, Geneva.

WHO (2014) WHO indoor air quality guidelines: household fuel combustion. World Health Organization, Geneva.

Wholey J. Evaluability assessment: developing program theory. *New Directions for Evaluation*, 1987, 33:77-92.

Wiltsey Stirman Kimberly J, Cook N, Calloway A, Castro F, Charns M (2012) The sustainability of new programs and innovations: a review of the empirical literature and recommendations for future research. Implementation Science 2012, 7:17

Woolcock, M., 2013. Using case studies to explore the external validity of 'complex' development interventions. *Evaluation,* 19(3), pp.229–248.

Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R (2013a) RAMESES publication standards: meta-narrative reviews. *BMC Medicine* 11:20.

Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R (2013b) RAMESES publication standards: realist syntheses. *BMC Medicine* 11:21.

Contact us

Centre of Excellence for Development Impact and Learning
London International Development Centre
36 Gordon Square
WC1H 0PD
www.cedilprogramme.org
@CEDIL2017