Approaches to evidence synthesis in international development: a research agenda

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

[1] Evidence for Policy and Practice, Information and Coordination Centre, University College London

[2] Department of Social and Policy Sciences, University of Bath.

ABSTRACT

This paper discusses the spectrum of synthesis methods available to generate, explore and text theory, their value to the field of international development and innovations required to make better use of the primary research available. 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.

Acknowledgements

Funding

Approaches to evidence synthesis in international development: a research agenda

1. INTRODUCTION

This paper aims to stimulate discussion about how evidence synthesis can be performed better.[1] 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. Conversely it may be viewed from the perspective of actual or 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 rigorous research methods (Gough et al 2017).

Evidence synthesis informs us of what is known from research, making it fundamental for informing policy decisions about development and for promoting the uptake and use of evidence from impact evaluations. The argument is beguilingly simple. Drawing on 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 to take into account contextual factors and values apparent in studies from different locations and standpoints.

Undertaking such systematic reviews provides timely access to the research evidence base that policy makers can use to inform policy development (in conjunction with other forms of

---

[1] A longer version appears on the CEDIL website (citation to be added).

evidence and other factors influencing decision making). A second purpose of evidence synthesis is to collate what we reliably know from prior research addressing development impact before embarking on new research.

Complex problems and choice of questions

Priorities for international development, identified either politically (Sustainable Development Goals, G20 Summit 2017, HM Treasury and DFID 2015) or by wider consensus (The Sheffield Institute for International Development 2015), raise questions about: global peace, security and governance; resilience and response to crises; global prosperity, extreme poverty and inequality; population dynamics, migration and refugee flows; and climate change. Cutting across these priorities are questions about impact, scaling up and delivering value for money (DFID 2015, 2016).

These priorities offer a spectrum of research problems or questions from relatively 'simple', through 'complicated' to 'complex'. In the field of international development choices between tackling simple-but-superficial and profound-but-wicked problems abound: a problem highlighted by the development dilemma that 'the more measurable the action the less likely it is to be transformative' (Natsios 2010). 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. Systematic reviewers have the option to recognise the inherent complexity in all interventions and their application or, conversely, to ask and address simple questions about complex problems (Petticrew et al 2015). For instance, the complex problem of day care 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 day care; perhaps create typologies of day care, analyzing the barriers and facilitators of effective care, and how different elements in the system interact' and more.

Complex problems and choice of research methods

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 of loss of credibility. 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.

Methods of interest here are those that render systematic reviews relevant to policy (and policy makers) by presenting 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.

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

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 cook stoves for cleaner air and accident prevention in homes (Puzzolo et al 2013).

To achieve this range of systematic reviews, methods vary in their attempts to generate, explore or test 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. This spectrum of approaches is illustrated in Figure 1 and the synthesis methods currently available across this spectrum are described elsewhere (Gough et al 2017).
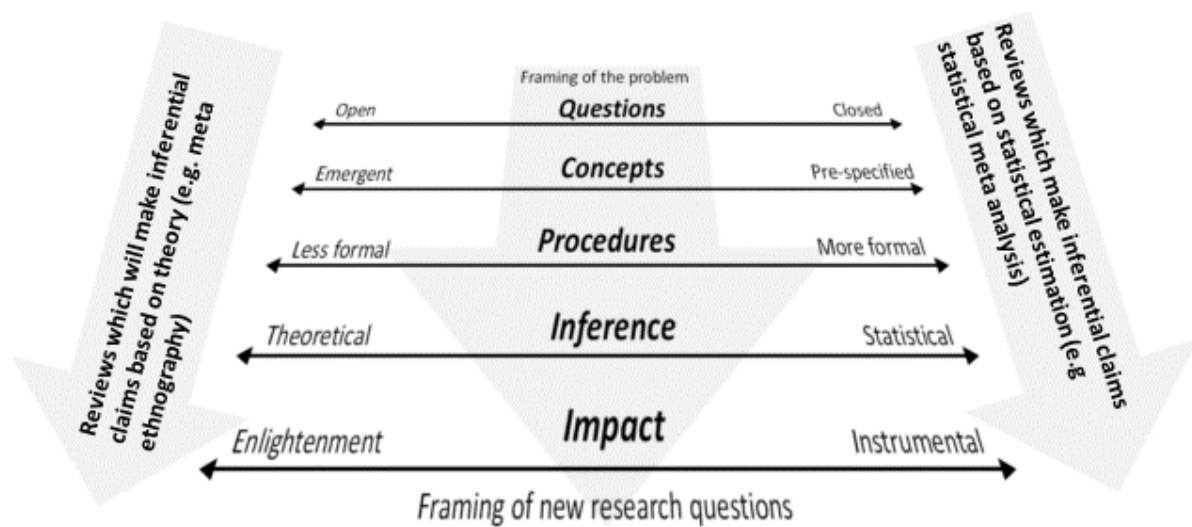
Figure 1: Dimensions of difference' in approaches to systematic reviews (Figure 3.11 in Gough et al 2017)

The next section considers how well existing methods match the field of international development. This is followed by the latest innovations emerging or required and a research agenda for advancing synthesis methods.

2. APPLYING EVIDENCE SYNTHESIS METHODS TO INTERNATIONAL DEVELOPMENT

In 2010, the Department for International Development (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 describing research efforts and synthesising research findings;
- 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 policy makers into narrower, manageable questions for synthesis; emphasising programme theory to structure a synthesis; and including a broader range of literature, some if 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.

Matching synthesis methods and innovation to the context of policy making

Systematic reviews vary not only in terms of the clarity of key concepts and theory before they begin, but also in terms of: whether they are produced either as a 'public good' for a general audience because problems are widespread and enduring, or as a product tailored with specific concerns in mind; and whether prior consensus on key concepts and definitions is strong or weak (Oliver and Dickson 2016). Figure 2 combines these in a two-by-two matrix.
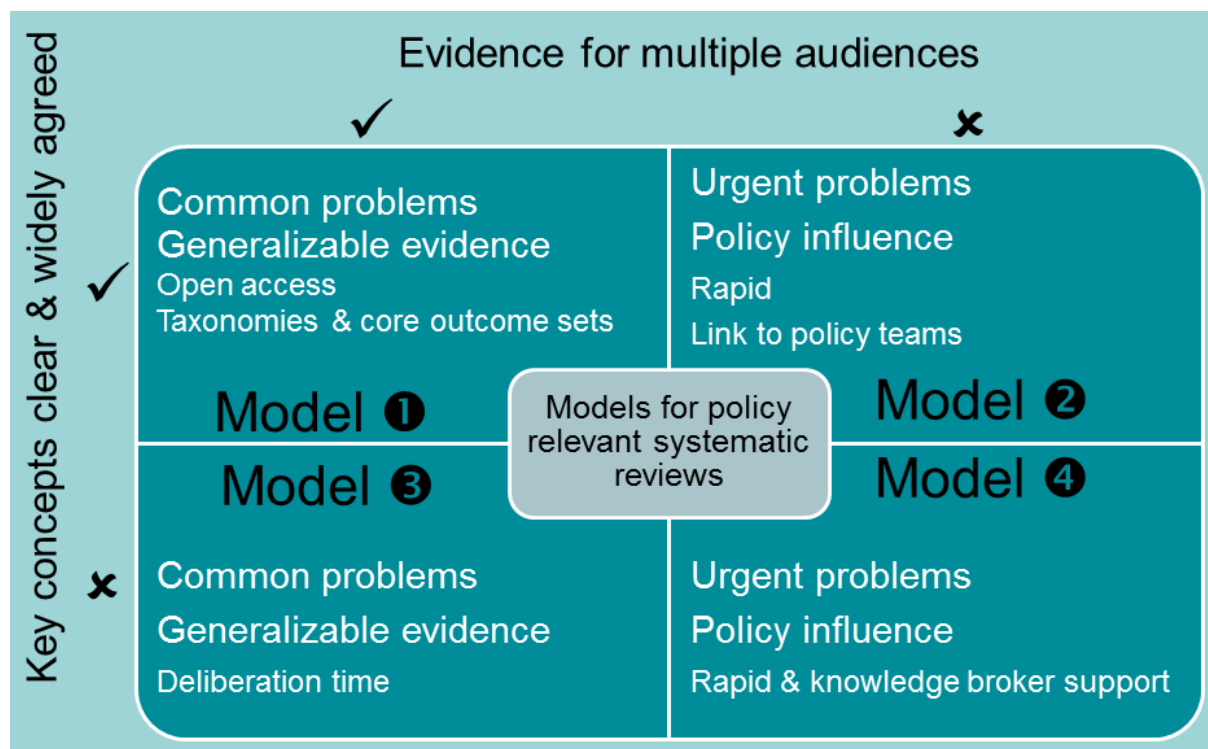

Figure 2: Four models of policy-relevant systematic reviews (adapted from ADD)

Model one, is applicable when there is 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 policy makers 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 et al 2017).

These can models 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-analysising '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 policy makers. Novice systematic reviewers require longer to complete individual reviews and need to accrue experience before applying state of the art methods with 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 policy makers 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.

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.

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 have more potential to serve vertical programming (eg HIV/AIDS, microfinance) than horizontal programming (eg 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).

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 homogenous interventions for homogenous populations where 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, policy makers 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, policy makers may ask what interventions work when outcomes of interest are broad, or vaguely unspecified (see Box 1 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 1: 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 of 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 in 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 proved 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 synthesize such research and identify the necessary future primary research to help further develop causal theory.

In this field, policy makers' international remits prompted the NGO 3ie[2] 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). Policy makers' 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).

3. 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 develops methods that:

- are well developed in other academic disciplines but rarely applied to international development;
- take advantage of developments in information technology for automation;
- 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;
- 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

---

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

- support asking better questions, not just developing more reliable answers; and
- accrue learning from studies addressing how evidence is used.

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 synthesizing evidence, and using synthesized evidence to support practice and policy across disciplines. GESI will achieve these aims through supporting Evidence Synthesis Centers based in LMICs.

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[3]) 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 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

---

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

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, stigmatization 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.

Automation

For some time synthesis methods have been taking advantage of emerging developments in automation. The field of healthcare is leading the way, thanks to a longer history of investment in better research curation than other fields; these new developments have wide applicability however. The potential of automation can be considered in two broad areas: making parts of the existing systematic review process more efficient; and changing the way in which research evidence is organised in more 'upstream' ways

In terms of making existing methods more efficient, new technologies are being developed which assist at each stage of the review process, with those assisting in the earlier stages more mature than those in the later parts. When developing search strategies, text mining and natural language processing technologies can be helpful in increasing both the precision and sensitivity of searches and in 'translating' search strategies across different data sources (Stansfield et al 2017). Given the time-consuming and repetitive nature of manual citation 'screening' (the activity of sifting through the – often – thousands of irrelevant research records in order to find the relatively few that match the review's inclusion criteria), much effort has gone into automating the identification of studies and reducing the human resource required for this activity. Two main approaches are emerging: an iterative workflow whereby a machine learning model progressively 'learns' the review's scope and suggests potentially relevant research for manual checking (O'Mara-Eves A et al 2015; Thomas 2013); and the development of classifiers which can recognise key characteristics of a field of research (e.g. distinguish between reports of randomised controlled trials and other research designs), and thereby eliminate large numbers of reports with high reliability (Thomas et al 2017). The latter stages of the review process are also receiving attention with development of tools for automating the 'mapping' of research, the assessment of bias in included studies, and even synthesis[4] (Stansfield et al 2013; Marshall et al 2015).

In addition to making aspects of the systematic review process more efficient, other developments are concerned with the upstream identification and classification of research, outside of the context of any specific reviews. Here, the 'evidence surveillance' concept is gaining ground in which the prospective and continual scanning of relevant sources (e.g. bibliographic databases, registries of trials, and grey literature from web searches) seeks to identify all research in a broad area and classify it in terms of the review questions it might address. Such systems aim to change the way in which reviews are conducted and updated and major producers of systematic reviews in health – including Cochrane and the UK's National Institute for Health and Care Excellence (NICE) – are investing to develop surveillance infrastructure and processes.

---

[4] RevMan-Hal: schizophrenia.cochrane.org/revman-hal-v4

While many of the above tools are being developed with a focus on health, many have widespread applicability – including support for automating (and semi-automating) aspects of the systematic review process. The IT infrastructure built for 'evidence surveillance' systems can also be utilized broadly, though the significant quantities of data on which machine learning depends may be less available in some areas. In broad terms, the technologies which automate study identification and evidence surveillance are likely to be more suitable for use in review models 1 and 2 (Figure 2); and tools which 'map' research automatically and learn iteratively from review decisions useable in all four review models. Infrastructure to support the use of automation in evidence synthesis requires significant investment which goes beyond the capacity of individual reviewers to build for themselves. There is therefore a need to consider at the strategic level how the field of international development might benefit from the use of these tools. Required work is likely to include research which evaluates how specific technologies can help, and the development of data repositories which provide: 'training' data for machine learning; and material for reviewers seeking research for inclusion in systematic reviews.

Accounting for differences in the context of primary studies

Scaling up interventions or introducing them into new 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).

This challenge is currently addressed in systematic reviews by reporting guidelines for addressing population differences, specifically in terms of equity (Welch et al 2012; O'Neill et al 2014). Although these guidelines provide a framework for epidemiological analyses well suited to public health, they are based on the PROGRESS tool, a mnemonic for capturing specific determinants of health: place of residence, race/ethnicity, occupation, gender, religion, education, socio-economic status and socio-economic position). In terms of the umbrella model of social determinants of health, this tool 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.

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 accommodate qualitative and mixed methods research designs have complemented this evidence of impact with findings to explain variations and disappointing impact. 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 OEDC'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

13

evidence to be re-cast as a synthesis for developing interventions. This creates an opportunity for systematic reviews to inform adaptive management by drawing on emerging findings from multiple examples of evolving programmes.

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 (Oliver et al in press).

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 3. It recognises the value of involving various stakeholders from policy, administration or delivery organisations and populations where impact is sought (right hand column of figure 3) in guiding systematic reviews. These stakeholders have the potential to influence review production (middle column of figure 3) 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).
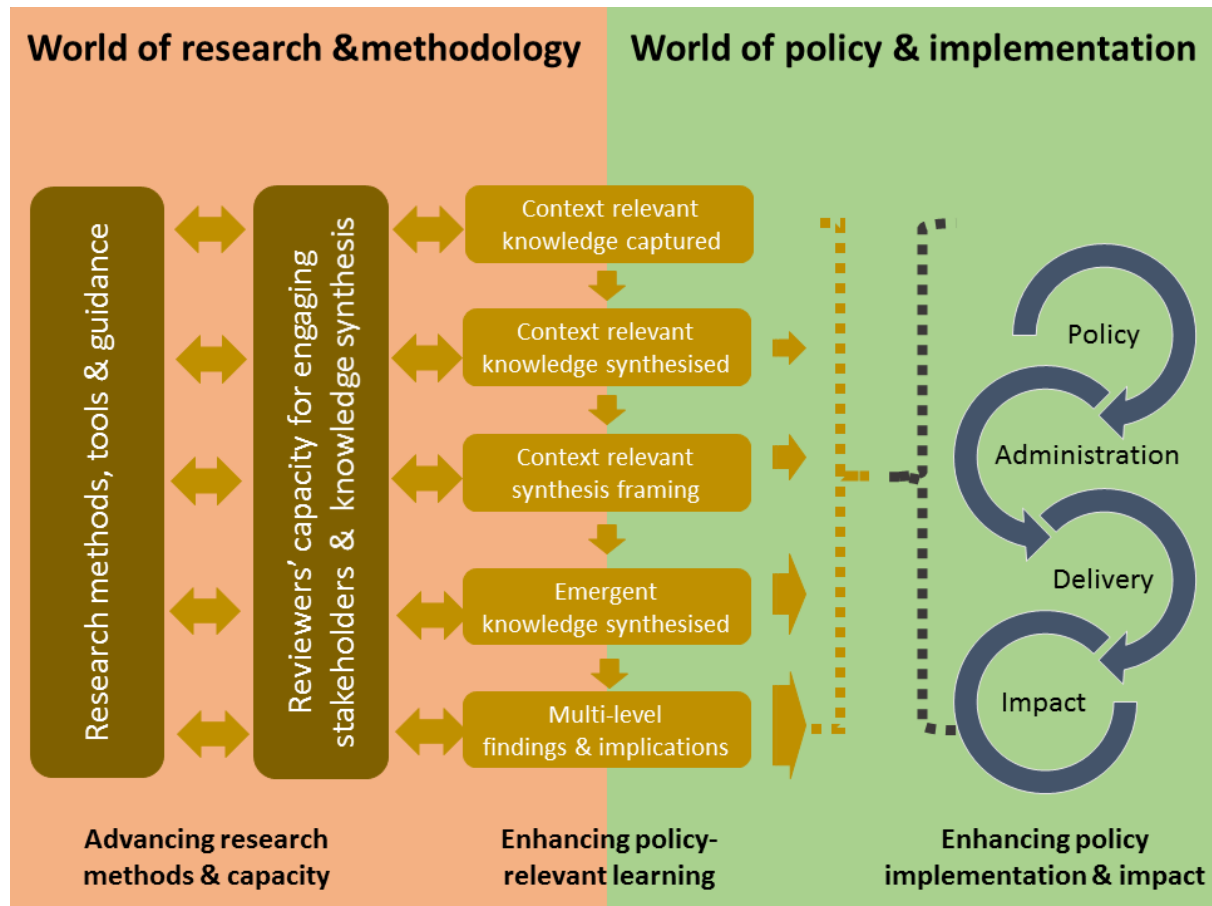

Figure 3: Potential advances in synthesis to enhance impact in the 'real world' (adapted from Oliver, Garner and Heywood (2015), unpublished discussions)

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

15

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).

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.

Asking better questions

Systematic reviews in international development have been driven more by policy makers than by academics, with policy makers asking questions which require a broader sweep of the literature. One solution has been two-stage reviews that first describe the nature and scale of the research available and then, before synthesising the findings, considers reducing the scope taking into account the most important sub-questions and the type of studies available likely to provide reliable answers. 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 a knowledge brokers (Campbell et al 2011), for tailored reviews rather than public goods reviews. Involving broader groups of policy makers 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 et al 2017).

4. CONCLUSION

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.

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; and 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.

Capitalising on advances in automation technologies: Benefits here will come from adopting and adapting methods, and strategic investment in information infrastructure.

Addressing contextual factors and inequalities in systematic reviews: We recommend systematic reviewers explore the utility of inequalities frameworks (eg 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 on-going 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.

Developing methods to address sustainability of interventions and outcomes: This may start with exploring the applicability of existing sustainability frameworks to synthesis methods.

Framing questions for public goods systematic reviews: There is a need to develop methods that involve collaborative groups of policy makers and evidence providers to discuss theories of change and how to generate better questions from them.

Filling thematic and geographic evidence gaps: Developing methods in these ways will advance the synthesis of evidence for the development of interventions, and scaling up of those proved effective; investigating inequalities, particularly gender inequality; and research addressing 'the toughest challenges in the hardest places' (DFID 2016). Doing so will go some way to filling substantive evidence gaps in the areas of governance and infrastructure, conflict and humanitarian settings, and climate change.

References

Aarons, G.A., M. Hurlburt, S. McCue Horwitz. 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, E.A., R. Fadlalla, L. Ghandour, O. Kdouh, E. Langlois, J.N. Lavis, H. Schünemann, F. El-Jardali. 2017. The SPARK Tool for Prioritizing Questions for Systematic Reviews in Health Policy and Systems Research: Development and initial Validation. Health Research Policy and Systems 15: 77

Anderson, E., M. A. Jalles d'Orey, M. Duvendack, L. Esposito. 2016a. 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.

Anderson E., M.A. Jalles d'Orey, M. Duvendack, L. Esposito. 2016b. 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

Ang L. 2014. *Early Childhood and Peace Building in the Asia-Pacific Region: A Literature Review to Inform a Regional Research Agenda*, UNICEF.

Ang L., S. Oliver. 2015. *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.

Annamalai T. R., G. Devkar, A. Mahalingam, S. Benjamin, S.C. Rajan, A. Deep. 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.

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

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

Bronfenbrenner, U. 1995. '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.

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

Choi, B. C., A. W. 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, B. C., A. W. 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, P.H. Intersectionality's Definitional Dilemmas. *Annual Review of Sociology* 2015; 41:1–20.

Coutts, C., M. Hahn. 2015. Green Infrastructure, Ecosystem Services, and Human. *Health Int. J. Environ. Res. Public Health* 12(8), 9768-9798; doi:10.3390/ijerph120809768

Craig, P., P. Dieppe, S. Macintyre, S. Michie, I. Nazareth, M. Pettricrew. 2008. Developing and evaluating complex interventions: The new Medical Research Council guidance. *British Medical Journal 337 doi:* *https://doi.org/10.1136/bmj.a1655*

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.

Duvendack, M., J.G. Hombrados, R. Palmer-Jones, H. Waddington. 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., M. Ehren, M. Bangpan, M. Khatwa, F. Perrone. 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.

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

Lavis, J. 2009. How Can We Support the Use of Systematic Reviews in Policymaking? Plos Medicine Nov; 6(11): e1000141.

Fischer, K., E. Ekener-Petersen, L. Rydhmer, K. Edvardsson Björnberg. 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.

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

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

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)

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

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

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

Kneale, D., J. Thomas, K. Harris. 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.

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

Langer, L., R. Stewart. 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.

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

Lewin, S., C. Glenton, H. Munthe-Kaas, B. Carlsen, C. J. Colvin, M. Gülmezoglu, J. Noyes, A. Booth, R. Garside, A. Rashidian. 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.

Liu, H., J. Muhunthan, A. Hayek, M. Hackett, T.L. Laba, D. Peiris, S. Jan. 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.

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

Marshall, I., J. Kuiper, B.C. Wallace. 2015. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. 1–10.

Moat, K.A., J.N. Lavis, S.J. Clancy, F. El-Jardali, T. Pantoja 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.

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

Nasser, M., V. Welch, E. Ueffing, S. Crowe, S. Oliver, R. Carlo. 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, F.K., A.K. Hurtig, C. Ahlm, B.M. Ahlberg. 2012. "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* 12 (433) **DOI:** 10.1186/1472-6963-12-433.

Nowotny, Helga; Nowotny, H., P. Scott, M. Gibbons. 2001. *Rethinking science: knowledge in an age of uncertainty*. Cambridge: Polity.

Obuku, E.A., R. Stewart, R. Mijumbi, M. Ochan, F. Achana, D. Akena, A. Nakitende, A. Ssemata, A. Kinengyere, D. Semakula, R. Ssenono, A. Nsangi, R. Lalitha, E. Mwesiga, J. Akite, R. Basaza, W. Newbrander, J. Okello, N. Sewankambo, with K. Dickson, S. Oliver. .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, S., M. Bangpan, C. Stansfield, R. Stewart. 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., K. Dickson. .2016. Policy-relevant systematic reviews to strengthen health systems: models and mechanisms to support their production. *Evidence and Policy.* 12 (2) 235-259.

Oliver, S., M. Bangpan, K. Dickson. 2017. Producing policy relevant systematic reviews: navigating the policy-research interface.
https://doi.org/10.1332/174426417X14987303892442

O'Mara-Eves, A., J. Thomas, J. McNaught, M. Miwa, S. Ananiadou. 2015. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Syst Rev. 4(1):5.

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

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

Peters, M.D.J., C. Lockwood, Z. Munn, S. Moola, R.K. Mishra. 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.

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

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

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

Puzzolo, E., D. Stanistreet, D. Pope, N. Bruce, E. Rehfuess. 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.

Rose, P., M. Battock. 2012. *Review of the DFID Systematic Review Programme*. DFID, London, UK.

Sinclair, D., B. Zani, S. Donegan, P. Olliaro, P. Garner. 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.

Stansfield, C., A. O'Mara-Eves, J. Thomas. 2017. Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. Res Synth Methods:1–11.

Stansfield, C., J. Thomas, J. Kavanagh. 2013. 'Clustering' documents automatically to support scoping reviews of research: a case study. Res Synth Methods 4(3) 1759-2887

Thomas J. 2013. Diffusion of innovation in systematic review methodology: Why is study selection not yet assisted by automation? OA Evidence-Based Med. 1(2):1–6.

Thomas, J., A. Noel-Storr, I. Marshall, B. Wallace, S. McDonald, C. Mavergames, et al. 2017. Living systematic reviews: 2. Combining human and machine effort. J Clin Epidemiol Available from: http://www.sciencedirect.com/science/article/pii/S0895435617306042

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

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

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

Wang, S., J. R. Moss, J.E. 2005. 'Applicability and transferability of interventions in evidence-based public health', *Health Promotion International* 21 (1): 76-83 2006

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

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

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

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