Causal Chain Analysis in Systematic Reviews of International Development Interventions

Dylan Kneale[1], James Thomas[1], Mukdarut Bangpan[1], Hugh Waddington[2], David Gough[1]

[1]Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre), Social Science Research Unit, Department for Social Science, UCL Institute of Education, University College London

[2]International Initiative for Impact Evaluation (3ie)

Suggested Citation: Kneale D, Thomas J, Bangpan M, Waddington H, Gough D, 2018 Causal Chain Analysis in Systematic Reviews of International Development Interventions CEDIL Inception Paper 4: London

About CEDIL: The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by UKAID through DFID. The mission of the centre is to develop and promote new impact evaluation methods in international development.

Corresponding Author: Dr Dylan Kneale, email: d.kneale@ucl.ac.uk

Table of Contents

Abstract

Understanding the extent to which an intervention 'works' can provide compelling evidence to decision-makers, although without an accompanying explanation of how an intervention works, this evidence can be difficult to apply in other settings, ultimately impeding its usefulness in making judicious and evidence-informed decisions. In this paper, we describe causal chain analysis as involving the development of a logic model, which outlines graphically a hypothesis of *how* an intervention leads to a change in an outcome. This logic model is then used to anchor subsequent decisions in the systematic review process, including decisions on synthesis. In this paper, we outline the steps taken in building a logic model, which usually consists of a series of boxes depicting intervention components and processes, outputs, and outcomes with arrows depicting connecting relationships. The nature of these connecting relationships and their basis in causality are considered, through a focus on complex causal relationships and the way in which contextual factors about the intervention setting or population may moderate these. We also explore the way in which specific combinations of intervention components may lead to successful interventions. Evidence synthesis techniques are discussed in the context of causal chain analysis, and their usefulness in exploring different parts of the causal chain or different types of relationship. The approaches outlined in this paper aim to assist systematic reviewers in producing findings that are useful to decision-makers and practitioners, and in turn, help to confirm existing theories or develop entirely new ways of understanding how interventions effect change.

Glossary

Note: this glossary does not include a definition of causality. Instead, the text provides describes different standpoints that can be taken in identifying relationships as causal relationships (see section on a plurality of approaches to causality (and evidence)).

Causal chain analysis (CCA): involving the development of a logic model and its use to anchor subsequent analysis, aiming to provide empirical evidence for parts of the causal chain and information about contextual modifiers.

Complex intervention: Interventions dependent on the creation of complex causal relationships, which are non-linear and may lead to emergent outcomes.

Complicated intervention: Interventions dependent on a large number of components or stakeholders, although may not involve developing complex relationships.

Counterfactual (reasoning): Consideration of the likely outcome in the absence of an intervention.

Emergent outcomes: Emergent outcomes are those outcomes that only occur through the interaction of intervention components/processes (and populations and settings), with no individual component directly associated with the occurrence of the outcome (i.e. the whole intervention leads to an outcome that individual component parts cannot lead to). Consequently, our understanding of emergent outcomes may only develop as our understanding of the intervention itself advances.

Epistemology: The examination of the nature of human knowledge and the contribution of concepts such as belief, truth, knowledge and evidence to our way of 'knowing'. In the context of Causal Chain Analysis of systematic reviews, developing an **epistemological standpoint** is understanding how we 'know' whether relationships are causal; this involves developing an awareness of the type of causal reasoning we employ to identify these relationships and the type of evidence we draw upon to support this reasoning.

Framework synthesis: A matrix-based method synthesis technique involving the construction of thematic categories into which data can be coded, which can be defined deductively or inductively.

INUS condition: (INUS – insufficient but non-redundant parts of a condition which is itself unnecessary but sufficient for the occurrence of the effect) INUS conditions form parts of a configuration of multiple conditions that trigger an outcome, although the configuration may not be necessary to trigger an outcome; similarly, an INUS condition may not be sufficient to trigger an outcome by itself.

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

Mechanism: Mechanisms, or pathways of action, describe the nature of the action occurring between intervention inputs and outputs/outcomes.

Mediator: Mediators are those factors that lie on the causal pathway between the intervention and outcome.

Meta-analysis: The quantitative synthesis of quantitative evidence from studies, usually in the form of an effect size indicating the magnitude and direction of effect, as well as a measure of its precision.

Moderator: Moderators are factors that can amplify or dampen the relationship between exposure to the intervention and the outcomes; they may interrupt or support the causal chain but are not integral links of the causal chain (unlike mediating factors).

Necessity/Necessary causal relationships: Signify that an outcome cannot be triggered in the absence of a condition, for example, an intervention component or contextual or participant characteristic.

Network Meta-analysis: An extension of standard meta-analysis that facilitates the comparison of multiple intervention options that may or may not have been directly compared against one another within trial reports.

Qualitative comparative analysis: A method of synthesis that supports the identification of sufficient and INUS conditions in particular that lead to successful outcomes.

Realist synthesis: A type of systematic review that seeks to clarify and identify the theoretical assumptions of an intervention (in this case based on a logic model), and then inductively test these empirically in order to find configurations of: causal mechanisms (M) that lead to the desired outcomes (O) and can be traced back to identify relevant conditions (C).

Sufficiency/Sufficient causal relationships: signify that that an outcome is triggered in the presence of a sufficient condition or sufficient condition set, but that other pathways to achieving the outcome may also exist.

Theory based systematic review: The process by which a conceptual framework developed to represent an intervention is used to design all stages of the review, including the development of research questions and organisation of evidence.

Theory of change: As in the case of logic models, theories of change are used to graphically represent complex interventions. There is overlap although unlike logic models, theories of change are more explanatory as they require all of the underlying assumptions of how and why different components, activities and outputs lead to a change in outcomes to be hypothesised at the outset, as well as an indication of the context and the stakeholders affected; there can be multiple causal chains for different stakeholders.

Virtuous/vicious cycle: These are activated when initial changes in the outcome create the opportunities for further self-reinforcing changes.

Section 1

Introduction

1.1 Context of the paper

If we were to track the development of systematic reviews over time, we may observe that as our toolbox of analytical methods has expanded, so too has our ability to address questions involving the explanation of how interventions work, as well as if they work. This means that we have started to move from more simple accounts of causality to focus on alternative, invariably more complex, causal pathways that allow us to explain and sometimes predict intervention effects. Casual chain analysis describes an approach that uses different methods to theorise and test how interventions exert an influence over outcomes. This paper tracks some of this thinking, and an underlying argument that we make in this paper is that hypothesising how an intervention works at the start of the systematic review process, helps us to formulate and identify causal pathways, which can be tested using some of the synthesis methods outlined later on.

Logic models and theories of change are gaining prominence within systematic reviews across disciplines as a way of showing assumptions of how the intervention works in a pictorial form (Kneale et al., 2015). But in order to address questions of how interventions work, and to transform a 'picture' into a 'model' that supports all stages of the systematic review, it is important for systematic reviewers to think through the types of relationships that are being depicted, their relationship with populations and context, and the types of synthesis methods are most appropriate. This is especially pertinent in the field of international development interventions, where the types of intervention are often complex and context-sensitive, large scale, involving multiple components and stakeholders, and where the use of theory, in particular, can help to enhance the policy relevance of the evidence (Snilstveit et al., 2012).

Decision-makers are increasingly demanding that we tell a 'good story' in presenting evidence from systematic reviews of development interventions (Waddington et al., 2018). Logic models help reviewers to structure this story through supporting the development of relevant review questions, to structure evidence collection, and to present findings in a compelling way (Waddington et al., 2018). But more fundamentally logic models and subsequent causal chain analyses change the nature of the story of the intervention from a simplistic account of 'does it work' to a much more informative account of 'how it works and what happens for whom', where there is greater opportunity to engage with stakeholders in the development of the initial model, and greater opportunity to incorporate evidence that reflects the perspectives of communities and individuals into the synthesis (Oliver et al., 2018). The resulting story is relevant and useful for a wider

readership – policy-makers, commissioners, trialists, practitioners and researchers – and can enhance the relevance of the evidence for the populations they serve (Oliver et al., 2018).

1.2 Purpose of the paper

The overall aim of this paper is to develop the understanding of systematic reviewers synthesising evidence on development interventions who want to undertake a causal chain analysis to summarise evidence on how interventions work. This is ultimately in order to provide evidence that better meets the needs of a range of stakeholders.

In particular, the techniques described in this paper will aid systematic reviewers to:

(i) develop a logic model that incorporates an understanding of how different intervention components and processes effect change in outcomes, and the stages necessary to reach this change
(ii) think beyond a pictorial representation of an intervention and to hypothesise the nature of causal relationships being depicted within a logic model
(iii) to consider a number of different synthesis methods and approaches that support exploration of research questions that emerge from the development of a logic model
(iv) to identify elements of best practice in causal chain analyses

While mainly focussed on a systematic reviewing audience, decision-makers and practitioners will find several elements of this report useful through:

(i) showcasing a number of examples of causal chain analyses (CCA) in the international development literature that are intended to demonstrate the flexibility and benefits of a CCA approach to inspire future work;
(ii) identifying elements of best practice in the conduct of causal chain analyses that can be used in commissioning and assessing the quality of future reviews adopting CCA;
(iii) through providing recommendations for future commissioning streams that address a substantive and/or methodological gap in the conduct of causal chain analyses for systematic reviews.

Sections 2, 3 and 5 begin to explore how we think about causality. These sections focus on how we can use different causal accounts (Section 2) to develop causal explanations that can be assessed using different criteria and principles (Section 5), that may be reliant on identifying different types of causal relationships (Section 3).

Sections 4, 6, 7 and 8, are more applied in nature and focus directly on some of the tools and approaches that can be used for undertaking causal chain analysis in systematic reviews of international development interventions. Section 4 focuses on developing a logic

model and using this to guide causal chain analysis. Sections 6, 7 and 8 explore the utility of different synthesis approaches in systematic reviewing.

Section 9 provides a summary that directly contrasts the synthesis approaches described in earlier sections and the types of research question that can be addressed. This section also identifies principles of best practice for the conduct of causal chain analysis, while section 10 presents recommendations for CEDIL and DFID.

1.3 Why undertake causal chain analysis: The potential added value of thinking about mechanistic explanations of how interventions 'work'

Diarrhoeal diseases are the second most common cause of death in low-income countries, and understanding the evidence of how to reduce this burden is a major public health challenge. Improved sanitation that promotes safer disposal of human waste is thought to be an effective intervention. Two contrasting approaches to systematically reviewing evidence on the effectiveness of sanitation interventions can be found in reviews conducted by Clasen et al. (2010) and De Buck et al. (2017).

In the well-conducted review undertaken by Clasen et al. (2010), but one which did not employ a causal chain analysis approach, the outcome of interest was focussed on the incidence of diarrhoea. The review found evidence that interventions that promoted safe disposal of human waste were generally effective in reducing levels of diarrhoea, although there was heterogeneity in the magnitude of effect (and direction for one study). Because of difficulties in calculating appropriate standard errors for many of the studies due to their design, as well as other forms of conceptual and methodological heterogeneity, the studies were not formally combined in a meta-analysis. While the authors indicated that the review suggested that there was some evidence that the intervention was effective, they concluded that the review did not allow for the 'quantification' of the effect of the intervention. However, the absence of CCA meant the authors were not in a position to theorise whether the 'major differences among the studies, including the conditions in which they were conducted' were likely to mean that the intervention worked through a single causal pathway or multiple pathways. Likewise, through focussing on 'if' the intervention 'works', they were not directly able to make recommendations for future intervention design that an expanded focus on 'how' the intervention works is able to make.

A later review undertaken by (De Buck et al., 2017) did embody the principles of CCA, beginning with the development of a conceptual model (represented as a theory of change) of how different models of sanitation and handwashing interventions would lead to a reduction in mortality/morbidity. The focus of the quantitative synthesis was on more intermediate outcomes including the increased use of latrines and the reduction in open defecation. The quantitative synthesis was supplemented by implementation studies to understand how interventions were implemented and the relationship with contextual

factors. Among the results uncovered was that sanitation interventions appeared to be most effective when combined with handwashing interventions, although the longer-term impacts were less clear. One key enabling factor was the inclusion of community-based approaches in design and implementation of interventions. The authors attributed the use of their theory of change in the systematic review as being critical in terms of understanding the context in which the interventions were being implemented (De Buck et al., 2018). While there remained areas of ambiguity around the effectiveness (and particularly long-term effectiveness) of the intervention, as was the case for (Clasen et al., 2010), through endeavouring to undertake causal chain analysis, the authors were able to make a series of recommendations for decision-makers for policy, and crucially, were able to make recommendation for practitioners that could improve future intervention design (De Buck et al., 2017).

Both are examples of well-conducted reviews addressing different research questions, although the contrast in the breadth of findings above starts to illustrate the value of analysing the causal chain in reviews of complex interventions. The different ways of doing so are outlined in the remainder of this paper.

Section 2

Causal Thinking and Systematic Reviews

Section 2 Summary: Causal Thinking and Systematic Reviews

- This section discusses common types of reasoning that are used to identify causal relationships.
- Within the context of causal chain analyses of systematic reviews of international development interventions, systematic reviewers are likely to draw heavily on a mechanistic account of causality. However, reviewers are advised to draw upon different types of account in understanding how interventions 'work'.
- Mechanistic accounts aim to deconstruct causal relationships and to identify how an intervention channels an effect between intervention and outcome.
- Knowing the different properties of causal accounts is important in understanding the scope and certainty of the evidence claims that can be made and how evidence should be communicated to decision-makers.

Well-conducted systematic reviews begin with a clearly defined research question and an articulation of the conceptual framework (Gough et al., 2017, Davies, 2006, Waddington et al., 2012). In the context of systematic reviews of intervention studies, the conceptual framework is an articulation of how the intervention is expected to 'work' and to exert an impact on the target outcomes. A logic model provides a graphical representation of these assumptions (discussed in-depth below) through a series of boxes representing intervention processes, and outcomes linked by arrows indicating the direction of effect, which are developed into chains of cause-and-effect relationships (Rogers, 2000). But what do these arrows and boxes actually signify in scientific and philosophical terms; and exactly what kinds of relationships are being represented and with what kind of certainty? As discussed below, these depictions represent a number of ways of conceptualising causal relationships and different methods of establishing or identifying causal relationships (Cartwright, 2007b, Krieger and Davey Smith, 2016, Illari and Russo, 2014).

2.1 A Plurality of Approaches to Causality (and Evidence)

Relationships between an exposure and outcome can be defined as causal from a number of different epistemological standpoints and using a plurality of evidence (Krieger and Davey Smith, 2016). Reiss' review (2009) identified five main accounts and perspectives through which relationships are theorised as being causal in the social sciences, although there is substantial overlap between these. All five are discussed below, although are presented in the context of systematic reviews of development interventions:

(i) **Counterfactual accounts**, where we consider the outcome that would have occurred if an intervention had not been received. This has been described as 'a conditional with a false antecedent' (Illari and Russo, 2014), so for example 'in the absence of a microfinance intervention, there would be no added improvement in poverty levels' (this is not synonymous with assuming no absolute change in the absence of an intervention). This form of counterfactual reasoning is partly the basis for many common forms of impact evaluation methods (Gertler et al., 2016, Shadish et al., 2002) (see also below) and is also situated in some cases within broader 'difference-making' accounts of causality (Vandenbroucke et al., 2016).

(ii) **Probabilistic accounts** arise from statistical analyses of quantitative data (Reiss, 2009) and are important to reasoning about causality in social science (Illari and Russo, 2014). Many probabilistic accounts of causal relationships are based on classical linear regression models (Reiss, 2009), or extensions to these, and aim to model the effect of a ceteris paribus change (all other factors being equal) in one variable (intervention exposure) on another (outcome) (Wooldridge, 2015). Studies using observational methods, for example, cohort studies, use probabilistic accounts of causality, although relationships identified through observational studies are often undermined due to observed and unobserved

confounding factors. Probabilistic accounts of causality have been described as indeterministic or stochastic, in that they can indicate broad-brushed trends, for example at a population level, but random variation and observed and unobserved factors mean that they are not entirely deterministic. Probabilistic accounts of causality are important to consider in systematic reviews, as they can underlie the interpretation/extrapolation of evidence from randomised controlled trials (Cartwright, 2007b, Cartwright, 2010). The logic states that if the probability of a (desired) outcome occurring, for example, increase in vaccination rate or decrease in violence, given exposure to an intervention in a subpopulation (the treatment group) differs from a similar control group who were not exposed to the intervention, then the findings can also be extrapolated to the larger population that these groups represent (Cartwright, 2007b, Cartwright, 2010, Illari and Russo, 2014). However, this extrapolation can be problematic for a number of reasons; for example experimental and target populations may differ from one another in unobserved and unknown ways, or that the antecedents of outcomes may differ between experimental and target populations (for example Cartwright, 2007b, Cartwright, 2010, Krieger and Davey Smith, 2016, Reiss, 2009).

(iii) **Regularity accounts** identify causal relationships through successive observation of patterns to develop regularity theories of causation (Reiss, 2009). While these accounts can ostensibly appear to be some of the most 'minimalistic' accounts of causation, this type of causal account underpins some of the methods used to handle complexity in evidence synthesis. For example, synthesis techniques such as Qualitative Comparative Analysis are theoretically based on regularity accounts (Thomas et al., 2014, Cartwright, 2007b, Reiss, 2009), but are interpreted using mechanistic reasoning in systematic reviews.

(iv) **Mechanistic accounts** of causality aim to deconstruct causal relationships and to identify how an intervention channels an effect between intervention and outcome (Illari and Russo, 2014, Reiss, 2009). Logic models (described below) aim to develop a mechanistic theory of how an intervention exerts an effect on an outcome, through providing a framework for analysing intervention effects as causal chains. These causal chains outline the steps that need to occur in order to reach a particular outcome.

Mechanistic accounts aim to elucidate how entities (the components the intervention) and activities (what these entities do) are organised to effect a change (mechanism) in the outcome(s) (Illari and Russo, 2014). Mechanisms are often "middle-range" theory, lying between project level causal chains and grand theory of psychology, sociology or economics, but drawing on the both (Pawson and Tilley, 1997). These relationships can be highly context-dependent, and the

longer the causal chain, the more likely the greater the influence of context on these relationships (Krieger and Davey Smith, 2016, Rogers, 2000). In the case of international development interventions, failure to consider the influence of context on mechanisms (i.e. the external validity of the evidence and the extent to which the evidence can 'travel' to other places or people) can lead to unintended or harmful outcomes when interventions are transplanted from one context to another.

When using logic models (discussed below), these mechanisms may only be tentatively hypothesised at the start of the review, based more on logical reasoning than well-articulated theory or empirical data, and the review process itself provides evidence for the existence and nature of the mechanism (Kneale et al., 2015). While in principle all trialists should articulate the causal chain through which an intervention is expected to exert an effect on the outcome, in practice, these details can be surprisingly scant and it is often left to the systematic reviewer to describe the intervention and provide a mechanistic account of causality (Kneale et al., 2015, Maden et al., 2017). Developing such a mechanistic account of intervention causality also often draws upon evidence from other forms of causal account, e.g. counterfactual reasoning, in its creation[1].

[1] Although the converse is not true and there is no expectation that counterfactual reasoning involves producing a mechanistic account of causality.

The importance of thinking about mechanistic explanations of how interventions 'work'

'PlayPump' aimed to improve access to clean water by harnessing children's willingness to 'play' and installing a merry-go-round to pump water in place of a conventional water pump. This intervention had shown promise in some settings (Ika and Donnelly, 2017), but was ultimately unsuccessful when scaled up for a number of reasons (UNICEF, 2007, Ika and Donnelly, 2017). One critical factor was that in one of the countries in which pumps were newly installed, Zambia, stakeholders had not been consulted a priori, were not given any choice on installation and were actually more satisfied with their existing technology than the new 'PlayPumps' (UNICEF, 2007). In contrast in South Africa, where some of the earliest PlayPump sites were located, there were higher levels of accountability and engagement with local government structures and user communities. There was also a different water provision landscape at baseline, with Free Basic Water Policy operational in South Africa, although access to free water was not a universal policy across the other settings in which 'PlayPumps' were installed. In this example, theorised mechanisms, which rested heavily on children's willingness to play, were thwarted because there was little perceived need for the intervention compared to usual practice, and community engagement was deprioritised when transplanting the intervention. A mechanistic understanding of how the PlayPump improved access to water could have involved examining how the local context in South Africa supported the delivery of the intervention and would have explored the implementation of the intervention processes and its relationship with outcomes.

(v) **Interventionist accounts** of causality revolve around the notion that a causal relationship between exposure and outcome is something upon which we can imagine intervening upon to bring about change (Reiss, 2009, Vandenbroucke et al., 2016). Interventionist accounts of causality are implicit within systematic reviews of social interventions; i.e. at the basis of each trial, there is a focal 'antecedent' and an underlying belief that changing this antecedent will lead to a change in the outcome. However, interventionist accounts have been criticised as being 'ideal' and not 'real', in that they can overlook the fragility of relationships in the social world and the way in which outcomes can emerge through a number of different pathways. Similarly, interventionist accounts of causality can overlook the reality that a causal relationship between exposure and outcome may look very different from the causal relationship between intervention and a change in outcome (Kelly and Russo, 2017, Reiss, 2009); for example within public health, although the direct causes of obesity epidemic may be attributable to factors such as poor diet and a lack of physical exercise,

permanent changes in population levels of obesity may only be observed if other factors such as poverty become a joint focus. In the context of this particular paper, much of the evidence being reviewed may implicitly already adopt an interventionist perspective starting with the standpoint that complex social problems are malleable (to some extent).

Systematic reviews of international development interventions, which are by their nature complex interventions (see glossary for definition), may draw upon several of the lenses described above in conceptualising and identifying causal relationships, and the evidence that is synthesised is similarly pluralistic in order to address our research questions. In fact, drawing on a number of different approaches listed above is considered preferable because of deficiencies in the scope, coverage, or validity of any one of the accounts described above when used in isolation (Krieger and Davey Smith, 2016, Reiss, 2009, Reiss, 2012). Nevertheless, in the pursuit of examining 'how' interventions work, causal chain analyses draws strongly on mechanistic accounts of causality, although this will likely be supported by other forms of causal account (e.g. counterfactual reasoning). Developing causal arguments based upon different epistemological standpoints and the causal accounts described above helps to avoid privileging evidence drawn from certain traditions (i.e. quantitative) above others or employing a narrow definition of evidence, which has been a frequent criticism of systematic reviews in international development in the past (Cornish, 2015).

2.2 Epistemology of Causality in Systematic Reviews

For systematic reviewers, being aware of how we conceptualise and identify causal relationships, and how this influences our causal reasoning and choice of methods (Illari and Russo, 2014), forms our epistemological standpoint with relation to causality, which can represent a key 'dimension of difference' in the type of systematic review we are conducting (Gough and Thomas, 2017, Gough et al., 2012). Clearly, thinking through our epistemological standpoint in this way involves going beyond the quantitative and qualitative methodological divide that has been pervasive in social science (Gough et al., 2012). For example, systematic reviews employing quantitative synthesis methods (meta-analysis) may be drawing upon counterfactual reasoning of causal relationships but will also be drawing upon probabilistic accounts in their interpretation; while the act of synthesising effect sizes from different studies, particularly when exploring subgroup analyses, arguably also draws upon accounts of regularity to causal relationships where there is low heterogeneity within groups. Similarly, although the synthesis of evidence from qualitative studies of interventions may initially be conducted with a view of providing a mechanistic account of causality, reviewers may seek and identify patterns of regularity to aid their interpretation of causal relationships. Qualitative Comparative Analysis, for example, is identified as a method supporting a regularity account of causality (Reiss, 2009), but it is also employed jointly alongside meta-analysis in some systematic reviews

for providing mechanistic causal accounts of how effective interventions work (Thomas et al., 2014, Chandler et al., 2017, Brunton et al., 2015a, Ton et al., 2017).

Understanding our own epistemological standpoint around the types of causal accounts we are creating within a systematic review is perhaps most important when it comes to the types of causal claims we make from our reviews and how we want others to use our evidence. Cartwright distinguishes between methods for warranting causal claims that 'clinch' the conclusions, such as those based on probabilistic accounts of causality using statistical techniques, and those that 'merely vouch' for their conclusions, for example, QCA (Cartwright, 2007b). She highlights the weakness in terms of applicability of the former, and the uncertainty (and potential bias) surrounding the latter form of causal claim. Systematic reviews can arguably support elements of both types of claim, depending on the scope of the research question (or statement) and the methods employed. Furthermore, in many ways, systematic reviews might be considered an analytical method that can potentially strengthen both types of warrants for causal claims outlined by Cartwright (Cartwright, 2010), through different forms of triangulation employed during the conduct of a systematic review employing causal chain analysis (Hales, 2010).

For philosophers such as Illari and Russo (Illari and Russo, 2014), it is good practice to explain where one's theorising about causality stands with respect to epistemological and methodological standpoints. Given that systematic reviewers implicitly theorise about causality on a daily basis when synthesising evidence and making judgements on intervention effectiveness, setting out our epistemological stall with respect to causality should be common practice. The excess of 'bare bone' reviews, however (Snilstveit, 2012), characterised as lacking both a theoretical basis and policy relevance, suggest this is likely to be a rarefied practice. Nevertheless, a greater understanding of the type of causal account we are developing can help reviewers to understand the limits and warrants surrounding findings. While as a discipline, there has been a heavy focus on synthesis methods, and a focus on maximising internal validity, it is questionable whether this focus has been at the expense of a richer understanding of causality in epistemological and metaphysical terms. Increasingly, however, setting out an epistemological standpoint can happen more tacitly with the development of a causal chain model to anchor a review (Anderson et al., 2011, Kneale et al., 2015), and the identification of suitable synthesis methods to support exploration of the model. It is these analyses that form the basis of the remainder of this paper.

Section 3

Making Links Between Interventions and Outcomes

Section 3 Summary: Making Links Between Interventions and Outcomes

- This section explores how reviewers can conceptualise and identify mechanisms through which interventions change outcomes.
- In development interventions, mechanisms can be identified through theorising and making links between (i) the intervention component; (ii) the function or purpose of the component; (iii) the output or outcome it is intended to change; (iv) the type of causal relationship between component and outcome (and potential mediators and moderators)
- In complex interventions, mechanisms may take on non-linear forms, and different forms of complex mechanism are discussed.

A causal chain provides a summary of the sequence of activities and changes that link intervention inputs and desired outcomes. Building a causal chain involves identifying the entities (components of the intervention) and their activities (their behaviours or functions) and describing how these are organised and then channelled to effect a change in the target outcome. Together, these have been described as 'mechanisms' (Illari and Russo, 2014). From the perspective of a systematic review of an intervention, identifying a mechanism involves describing:

(i) The intervention component
(ii) The function or purpose of the component
(iii) The output or outcome it is intended to change
(iv) The type of causal relationship between component and outcome (and potential mediators and moderators) – i.e. how the effect is channelled

It is this latter feature that helps to distinguish between complicated intervention and complex intervention in terms of causality (Lewin et al., 2017, Rogers, 2008). For example, while interventions may involve a large number of components or stakeholders, and may therefore be complicated, they may not necessarily be dependent on complex causal relationships, which are non-linear and may lead to emergent outcomes (Rogers, 2008). Emergent outcomes are those outcomes that only occur through the interaction of

intervention components/processes (and populations and settings), with no individual component directly associated with the occurrence of the outcome (i.e. the whole intervention leads to an outcome that individual component parts cannot lead to). Consequently, our knowledge of emergent outcomes may only develop once we better understand the intervention itself (Rogers, 2008).

The most simple causal relationships are those where we assume (or test) whether the intervention has a linear effect, where a change in outcomes occurs after exposure to the intervention, and where greater exposure to the intervention is expected to be proportional to the impact. Often in the social world, these types of linear causal relationships can be difficult to substantiate, and we describe some of the more complex relationships below, which also form some of the building blocks of causal chain analyses.

3.1 Complex interventions and complex mechanisms

Glouberman and Zimmerman (2002) articulate the difference between simple problems (e.g. following a recipe), complicated problems (e.g. sending a rocket to the moon) and complex problems (e.g. raising a child). Determining whether an intervention should be regarded as simple, complicated or complex can lie in (i) the degree to which context is critical for the likely outcome achieved and hence ability to project generalisations from one case to the next; and (ii) more broadly the unpredictability of intervention effects and the extent to which the intervention can be viewed as an adaptive or learning system, evolving in response to the intervention (Chandler et al., 2017, Lewin et al., 2017). It is difficult to think of many if any, international development interventions that are 'simple' interventions that always give rise to an identical outcome when repeated time and again, and it is not clear to what extent many development interventions are complicated or complex in nature. However, Lewin and colleagues (Lewin et al., 2017) offer a tool for systematic reviewers to aid assessment of the degree to which an intervention can be regarded as more complex, focussing on the extent to which different dimensions of complexity - complexity in terms of components, implementation, context and participants - contribute to an intervention's impact. These dimensions of complexity give rise to the types of complex, causal, and non-linear relationships described below. All of the mechanisms described below can feature as parts of different accounts of causality laid out earlier.

Virtuous circles/cycles (and vicious circles/cycles): A 'virtuous circle' is activated when initial changes in the outcome create the opportunities for further self-reinforcing changes (Rogers, 2008). For example, a recent review on the mental health interventions and their impact on economic outcomes in low and middle-income countries concluded that 'improvements in economic status go hand in hand with improvements in clinical symptoms, creating a virtuous cycle of increasing returns' (Lund et al., 2011, p1502). In contrast, they found less evidence for virtuous circles operating in the reverse direction, where poverty reduction programmes did not appear to impact upon mental health outcomes. The converse, vicious cycles, are self-reinforcing negative intervention effects.

An example of a virtuous or vicious circle in international development is the relationship between economic growth and human development, as articulated in a causal chain with feedback loops in Ranis and colleagues (Ranis et al., 2000).

Tipping points and threshold/plateau effects: Tipping points occur when an intervention appears to have no discernible effect until a critical point has been reached (Shiell et al., 2008). Rogers (2008) also discusses tipping points in the context of virtuous circles and amplification, where a small amount of exposure to an intervention can have a disproportionately large impact on the outcome once a tipping point has been reached. Threshold effects have been described in a similar way, indicating the need for a critical value to be reached before an outcome is triggered. However, the notion of a plateau (or threshold), can also indicate a point of saturation where further change cannot be triggered within the confines of the context. For example, in review microfinance on women's control over household spending in developing countries, some studies described observing that a 'certain threshold level of independence within the structural norms of the society' had been reached and that 'microcredit has no [further] marginal impact on all such indicators' (Vaessen et al., 2014, p70). Another example is of a sanitation intervention that, due to environmental health spill-over effects, may only be effective when a threshold proportion of the community have access to, and use, the facility (see also, (Benjamin-Chung et al., 2015)).

Mediators, interaction effects and moderator effects: Mediators are those factors that lie on the causal pathway between the intervention and outcome. While mediators can be represented through linear causal relationships (i.e. not all mediators form parts of complex mechanisms), they are of interest as they can change the interpretation of causal chains. For example, in a systematic review underway on the effectiveness of interventions to raise children's educational and health outcomes through increasing women's empowerment, women's intra-household bargaining power and time use were identified as mediating factors (Vollmer et al., 2017). In other words, for the intervention to effect change in children's outcomes, it must also change women's intra-household bargaining power and time use. Most of the logic models and theory of change techniques discussed below have an explicit representation of mediators, although many meta-analytic models analyse these separately and not as part of a causal chain. In contrast, moderators and interaction effects refer to factors that can amplify or dampen the relationship between exposure to the intervention and the outcomes. While often represented as individual participant characteristics in program theory, in the absence of individual participant data in many systematic reviews, as well as the scarcity of stratified estimates in trial reports, these reflect study-level moderators in many meta-analyses (Kneale et al., under review). Examples include contextual factors such as poverty status, which might indicate an expectation of higher returns from a lower base, and factors relating to the quality of implementation which might predict the opposite where fidelity is low (Waddington et al., 2012).

Conjunctural causation refers to circumstances where a particular intervention component or contextual or participant characteristic triggers an outcome only in the presence of another component(s). **Multiple conjunctural causation** is an extension of this principle, which explores the possibility that the organisation of different sets, each consisting of different components/characteristics which alone cannot trigger an outcome, lead to the same outcome. Exploring causal relationships from this perspective involves focussing on the organisation of the constituent parts of mechanisms, and less on the way in which causal relationships channel their action.

Necessary causal relationships signify that an outcome cannot be triggered in the absence of a condition (a factor or variable), for example, an intervention component or contextual or participant characteristic. Necessary causal relationships can be based upon one condition, or a set of conditions (see conjunctural causation). An example of a potential necessary relationship can be drawn from the conclusions of a systematic review of interventions to prevent female genital mutilation/cutting (FGM/C) in Africa where, for example, in contexts where FGM/C is closely related to Islam, it is necessary to engage religious leaders in the intervention and develop a unified position among religious leaders, in order to ensure the program is successfully implemented (Berg and Denison, 2012); no examples of successful implementation were observed where religious leaders were not engaged with. A further example of a necessary condition can be observed from the conclusions of a recent literature review which suggests that computer/smartphone access is a necessary component of interventions that seek to enhance e-Government in sub-Saharan Africa, but is not sufficient to trigger this outcome without a legal framework that supports implementation also being in place (Nkohkwo and Islam, 2013). In this example, enhancement of e-Government cannot occur without computer/smartphone access (access is necessary), but may access alone is not sufficient to trigger the outcome (conjunctural causation).

Sufficient causal relationships signify an outcome is triggered in the presence of a sufficient condition or sufficient condition set, but that other pathways to achieving the outcome may also exist. These forms of sufficient causal relationships are usually the target of systematic reviews (Kristjansson et al., 2016, Stewart et al., 2015, Dangour et al., 2011).

INUS causal relationships (insufficient but non-redundant parts of a condition which is itself unnecessary but sufficient for the occurrence of the outcome) are an extension of the logic of sufficient and necessary conditions above. Mackie's (1965) classic example of an INUS causal relationship involves the role of a short circuit in starting a house fire, where a short circuit could only have triggered a fire in the presence of flammable materials nearby. A short-circuit alone is therefore not sufficient for a house to catch fire but in the presence of other components including flammable material (conjunctural causation), does become part of a set of conditions sufficient for causing a fire. However, this set of conditions is itself not necessary to start a house fire, as there are many other routes through which homes catch fire.

Some accounts of complex interventions expand on these and define complex interventions as those that share similar properties to the complexity of the wider systems in which they operate; complex interventions are composed of nested systems within a system which is itself complex (Clark, 2013, Shiell et al., 2008). This can help reviewers to conceptualise interventions and their focal point differently and can lead reviewers to consider the properties of a system that need to change in order for a change in outcome to be observed. This type of 'systems thinking' is becoming increasingly common within systematic reviewing (Noyes et al., 2013). Awareness of these different forms of causal relationship, as well as understanding the epistemological standpoints (see earlier section) allows us to take the first steps in undertaking causal chain analysis in systematic reviews, and that is to conceptualise the causal chain itself.

Section 4

How Can Causal Relationships be Developed into Causal Chains and Theories of the Way in Which Interventions Operate Within Systems?

Section 4 Summary: How Can Causal Relationships be Developed into Causal Chains and Theories of the Way in Which Interventions Operate Within Systems?

- Logic models are graphical representations of programme theory that depict intervention components, mechanisms (pathways of action), outputs, and outcomes as sequential chains of events. These form the basis of causal chain analysis.
- Logic models are usually developed starting with the identification of outcomes, and theorising the necessary pre-conditions to reach those outcomes. The steps and necessary pre-conditions work backwards until the intervention itself is represented (the full stages are described within the section).
- Logic models can be a useful tool throughout the review process and serve as a basis for causal chain analyses. Reviewers are encouraged to update logic models on the basis of new understandings about the intervention developed from the systematic review process.
- Logic models can vary in complexity. Reviews of development interventions are best supported by models that theorise processes occurring within the intervention system, as well as relationships between the intervention and the broader system, and how these may vary across settings.

"Theories of change" and "logic models" (see below and glossary for definitions) are forms of programme theory that depict intervention components, mechanisms (pathways of action), outputs, and outcomes graphically, represented as sequential chains of events, and form the basis of causal chain analysis (Wirtz, 2007). Programme theory can form an anchor to most major decisions taken within the systematic review process, from the scope of the inclusion and exclusion criteria, through to the synthesis and interpretation of evidence (Anderson et al., 2011, Kneale et al., 2015, Waddington et al., 2012). While the use, and particularly effective and extensive use, of these techniques within systematic reviews is still in its infancy (Kneale et al., 2015, Maden et al., 2017), systematic reviewers are

encouraged to include a logic model or theory of change from the protocol stage to aid in theorising how the intervention might work (Campbell Collaboration, 2017). From the perspective of accounts of causal relationships discussed earlier, effective use of programme theory is instrumental in developing mechanistic accounts of how interventions affect a change in outcomes.

4.1 Approaches to theorising how interventions work

The terms "theories of change" and "logic models" are often used interchangeably by reviewers, largely dependent on disciplinary preference (Kneale et al., 2015). Within the evaluation literature, however, a somewhat fuzzy distinction exists between logic models and theories of change. Theories of change are often used to denote complex interventions, particularly where assumptions of how and why program components effect change are pre-specified, and based on existing knowledge. Logic models on the other hand are used to outline program components and check whether they are plausible in relation to the outcomes; they do not necessitate all of the underlying assumptions and mechanisms to be stated a priori (Funnell and Rogers, 2011, Clark and Anderson, 2004). Similarly, not all details of the contexts and stakeholders who are likely to be affected will be outlined in full in a logic model, and some parts of the causal chain may be treated as a "black box" requiring further unpacking as part of the review (White, 2018, Kneale et al., 2015). This distinction may fit in well with the different stages of many systematic reviews. A logic model provides an early depiction of the components of interventions and their outcomes, but not necessarily an extensive articulation precondition that is needed to achieve these outcomes, knowledge of which may be developed during the process of conducting the systematic review. Within the methodological literature on systematic reviewing, 'logic model' has emerged as the favoured terminology and the preferred tool for depicting intervention causal chains, hence we use the term hereon to describe the depiction of causal chains in reviews.

New taxonomies and ways of viewing logic models are increasingly allowing for complexity to be incorporated into what were previously more linear forms of logic model (Davies, 2018). Rohwer and colleagues offer a distinction between systems-based (depicting the interaction between an intervention and the system in which it takes place) and process-based logic models (depicting a temporal sequence of events) (Rohwer et al., 2017). Another paper in this series also examines the way in which conceptual frameworks are used for different purposes across disciplines, contrasting differences between epidemiology and economics (Vigneri et al., 2018).

4.2 How to Develop a Logic Model for a Systematic Review as the Basis for Causal Chain Analysis

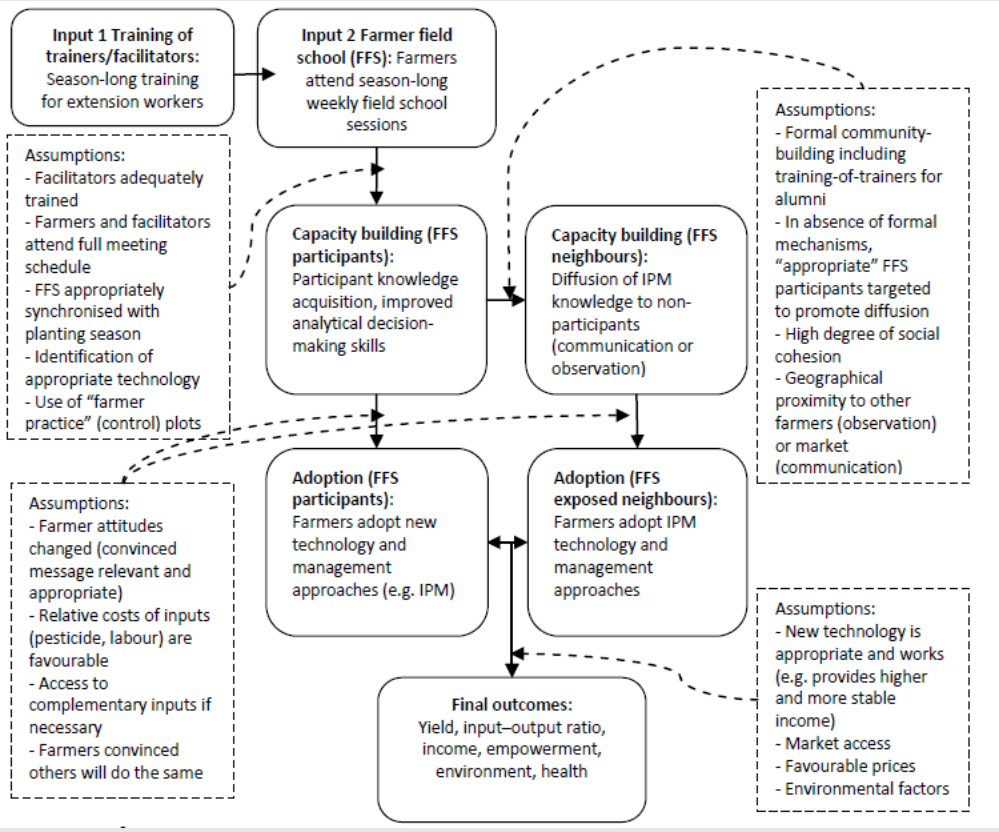
4.2.1 What Does a Logic Model Look Like?

Several examples of logic models exist in the systematic review literature (see (Kneale et al., 2015) for a snapshot review of those used in systematic reviews of international development interventions). The example below (figure 1), from a systematic review of farmer field schools to improve outcomes for farmers, was described as a 'hypothesised causal chain' (Waddington et al., 2014a, p33). Farmer field schools bring together groups of farmers in a neighbourhood together to learn about best practice in agriculture through participatory models of education, including field-based experiments on neighbouring plots of land through a growing season to examine the impact of best-practice techniques; historically the model had a focus on exemplifying the benefits Integrated Pest Management approaches, as opposed to pesticide spraying (Waddington et al., 2014a).

This model traces the way in which outcomes (e.g. yields) are hypothesised to be determined by the presence of intermediary conditions (adoption of new technologies among participants and diffusion effects among neighbouring farmers; for example Integrated Pest Management (IPM)); these are themselves shaped by a set of assumptions around supporting factors operating at a contextual level (for example market access). In turn, these adoption factors are themselves predicated on achieving a set of circumstances, reflecting capacity issues, which are again contingent on a set of contextual assumptions (Waddington et al., 2014a).

While the 'type' of complex causal mechanism (e.g. any hypothesised tipping points) is not directly stated for all connections, as is rarely the case in logic models, these can sometimes be expressed in footnotes to a logic model (Funnell and Rogers, 2011, Rogers, 2008), and the logic model itself could be used as a tool to help theorise the nature of these connections (Harris et al., 2015, Kneale et al., 2015). Furthermore, some of the complex causal relationships discussed in section 3 above are represented in Figure 1; for example, a virtuous cycle is depicted with adoption at a participant level leading to adoption by neighbouring farmers, and further reinforcing adoption by participants (Waddington et al., 2014a). In addition, in the farmer field schools systematic review, the authors returned to the theory of change in the final analysis drawing on the evidence synthesised, and articulated different causal chains relating to empowerment, this being an example of theory-building synthesis in the context of an effectiveness systematic review.

Figure 1: Logic Model for a Review of Farm Schools (taken directly from (Waddington et al., 2014a))



Notes: FFS = Farmer Field Schools; IPM = Integrated Pest Management

4.2.2 Steps in Building a Logic Model

The steps taken in developing a logic model afresh have been outlined in detail in Kneale and colleagues (Kneale et al., 2015), and are only briefly discussed here. Many other resources also exist to help trialists and reviewers to develop logic models, including well-known contributions by Funnell and Rogers (Funnell and Rogers, 2011), as well as more recent contributions focussed on systematic reviews (Anderson et al., 2011, Baxter et al., 2014, Pfadenhauer et al., 2016, Rohwer et al., 2017).

A starting point is for reviewers to familiarise themselves with the expected and intended outcomes of the intervention under study, and their potential mediating factors, as well as to consult existing logic models (or similar program theory techniques); program theory from related interventions may also be relevant to consider. Rohwer and colleagues provide two logic model templates, intended to provide a starting point for systematic reviewers, which may also be useful for reviewers starting from scratch (Rohwer et al., 2017). The causal chain is developed through the identification of distal/final outcomes,

and then the reviewers work backwards to identify or hypothesise the necessary preconditions (intermediate/ proximal/ mediating variables) to reach these distal outcomes. The ultimate aim is to create a chain of links between the intervention and the final or endpoint outcome. Several "links" could be added to the outcome chain, with a rule of thumb being the greater the complexity or length of the outcome chain, the more likely that the mechanisms may be influenced by or dependent on contextual factors (Krieger and Davey Smith, 2016). Intervention outputs can also be identified after identifying outcomes, those necessary pre-conditions to reach outcomes but not necessarily goals in themselves.

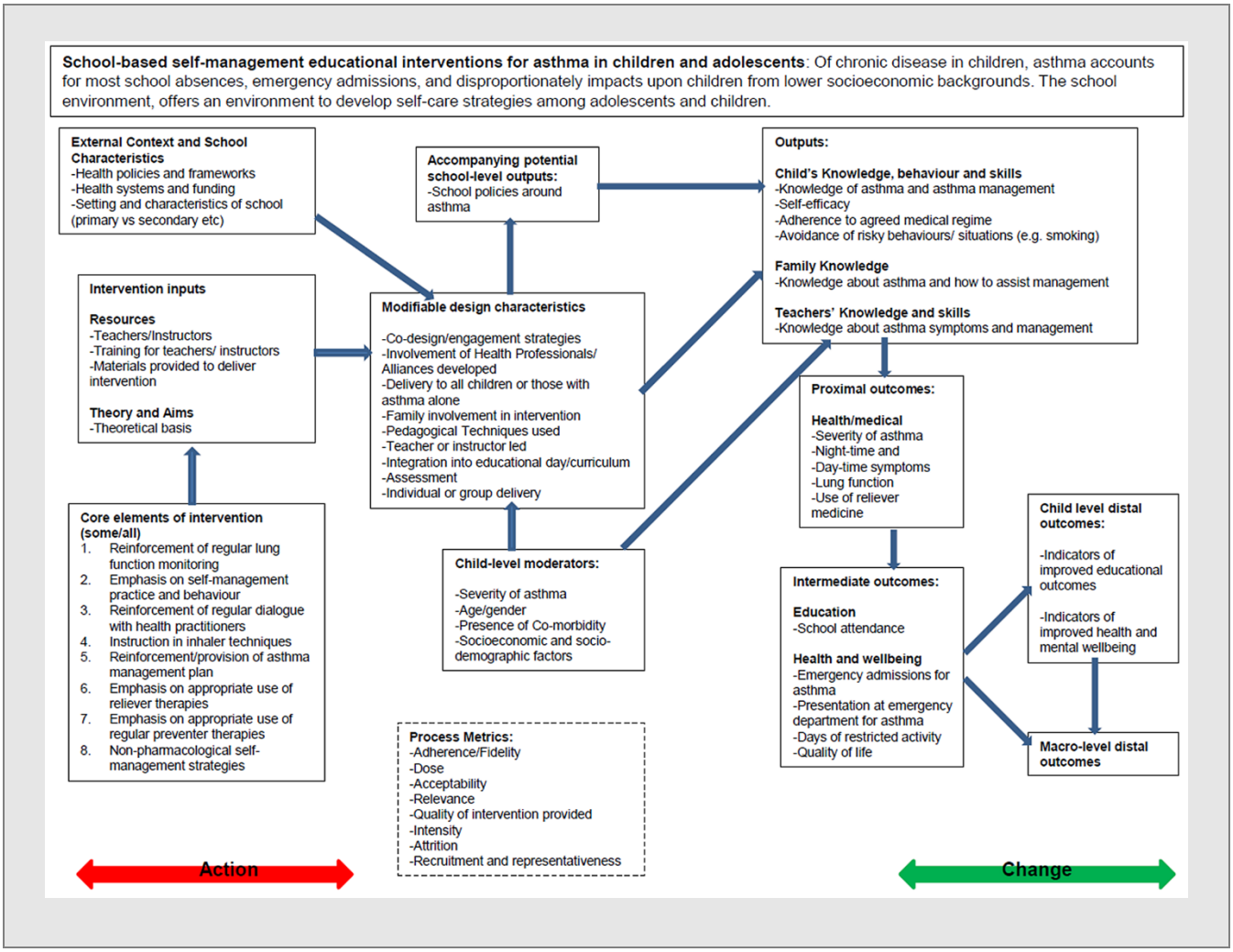
Continuing to work backwards from the outcome(s), intervention chains of intervention inputs are then specified. After completing input chains (composed of a programme's components, including any inception and implementation phases) and output and outcome chains, additional external or contextual factors can be theorised and represented as potential moderators. It is expected that several iterations of the logic model may be produced before a review team settles on a preferred model, with iterations representing an improvement in clarity, the conceptual soundness, and more logical sequencing and organisation of the causal chain. External stakeholders (lay members as well as trialists) can also be integral in forming a sound logic model (Rees and Oliver, 2012), and models of stakeholder engagement are examined in a different paper in this series (Oliver et al., 2018). Some logic models may explicitly identify areas of ambiguity (e.g. 'black box' of intervention or effects) where the synthesis contributes to understanding the causal chain. An example of a logic model developed through this process is displayed below for school-based asthma interventions (figure 2; see (Harris et al., 2015) for further information).

Finally, assumptions underlying the causal chain – those factors which determine whether the links in the chain follow one another – may be articulated (although extensive articulation of mechanisms is more commonly a feature of theories of change (Kneale et al., 2015)). These can include assumptions or targeted exploration of factors such as treatment take-up (White, 2018, Kneale et al., 2015). In White's paper on theory-based systematic reviews, a funnel of attrition outlining how the reach and impact of an intervention can diminish from the intended pool of beneficiaries is used as a heuristic device for thinking through the causal chain and its assumptions (White, 2018, p6). Assumptions may be inserted at any stage in the causal chain, between programme design and implementation, or outputs and outcomes, and so on (see example in figure 1).

Logic models may also be used to theorise unintended outcomes and potentially negative and harmful outcomes (Bonell et al., 2014). Causal chain analysis within evidence synthesis provides a method for providing mechanistic accounts of how interventions may deviate from their intended outcomes, a process described as modelling "dark logic" within interventions by Bonell and colleagues (Bonell et al., 2014). For example, a review of women's self-help groups in low and middle-income countries identified pathways towards economic and psychological empowerment resulting from the intervention but also

recognised potential alternative pathways towards adverse effects such as stigma or domestic violence (Brody et al., 2017).

Figure 2: Logic Model for a Review of School Based Asthma Interventions (see (Harris et al., 2015))



These steps outlined above are generally consistent regardless of the type of systematic review and form of causal chain analysis being undertaken, except for realist reviews, where there may be greater emphasis on formal identification of theory in the scoping stages (see later section on realist reviews and (Booth and Carroll, 2015b)). It is expected that the review process itself, through the emergence of new knowledge and findings, will lead to changes in a logic model, or an entirely new way of understanding how the intervention works, that can be used to help interpret and communicate findings (Kneale et al., 2015, Rehfuess et al., 2017). The a priori version of a logic model should be included in the protocol with details on how it will be used in later stages of the review.

Although the examples above tend to involve single, albeit long and complicated, causal chains; reflective of the complexity of International Development interventions, there is

scope for logic models to incorporate multiple simultaneous causal chains leading to the same, or different outcomes (Rogers, 2008, Funnell and Rogers, 2011). Similarly, there may also be a need to construct multiple logic models for large interventions to reflect the complexity of the intervention, or to guide multiple linked reviews.

Finally, software can support the development of logic models. While the examples above have been created through standard packages (Word or PowerPoint) having first been sketched out on paper, some may find specialist software advantageous. A recent example includes Dylomo, a free (at the time of writing) piece of software (Snow and Snow, 2017), which allows users to visualise changes in the intervention system; the authors also usefully review the features of eight alternative tools (free and subscription-based) that support the production of logic models. Some tools also exist that are more specialist to certain sub-disciplines within international development, such as Miradi (https://www.miradi.org/); specialist to conservation. Plans for future iterations of EPPI-Reviewer, specialist systematic review software (Thomas et al., 2010), also include provision for creating a logic model within the software.

4.3 Using a Logic Model as Part of the Systematic Review Process and in Causal Chain Analyses: Good practice and cautionary notes

In broad terms, logic models provide a framework for "thinking" conceptually before, during and at the end of the review (Anderson et al., 2011, Waddington et al., 2012). Within the review process, logic models can aid in (i) clarifying the scope of the review; (ii) identifying points of uncertainty that could become focal points of investigation; (iii) clarification of the scope of the study and particularly in distinguishing between different forms of intervention study design; (iv) ensuring that there is theoretical inclusivity at an early stage of the review; (v) clarifying inclusion and exclusion criteria; (vi) informing the search strategy with regards to the databases and scholarly disciplines upon which the review may draw literature; (vii) providing a communication tool and reference point when making decisions about the review design; and (viii) providing a project management tool in helping to identify dependencies within the review.

For causal chain analyses in systematic reviews, logic models provide an anchor for systematically investigating putative relationships in a causal chain (Wirtz, 2007), using some of the synthesis methods outlined here and elsewhere (Gough et al., 2017). Using logic models as a framework, pathways can be systematically decomposed into lower-level pathways (Wirtz, 2007), with the ultimate objective of identifying the most influential sub-chains and longer strands. Although a complete causal chain is rarely fully identified and measured in practice, a logic model provides the reviewer with the framework for theorising, explicating and empirically testing causal relationships and mechanisms within the causal chain.

Despite the utility of using a logic model to theorise how an intervention works and its role in supporting subsequent review decisions, it is useful to remember that a logic model is a

model and that evidence uncovered during the review process can challenge this model. We have described a logic model as providing an anchor to subsequent review decisions, although the initial model described may not be supported by the data and entirely new ways of thinking about the intervention may emerge from the evidence synthesised during the review. We reiterate that we would expect the logic model to develop on the basis of new understandings brought by systematic review (Kneale et al., 2015), and as such, there is a degree of flexibility in the initial theorising that can change as a result of new knowledge. This type of flexibility is perhaps essential when reviewing complex interventions, for example to account for emergent outcomes, which may only develop on the basis of a greater understanding of the intervention and how it works.

Rehfuess et al. (2017) take this notion of iteration further and identify an iterative approach to the development of a logic model, where the logic model is continuously updated throughout the review process. This flexibility in the logic model and its use for CCA does raise concerns about the introduction of potential bias. For example, Rehfuess et al. (2017) discuss the potential concerns around replicability and transparency in the review process that an iterative approach can introduce as well as the potential introduction of reporting bias (pathways that are not evidenced are omitted). Others may hold concerns around the introduction of confirmation bias, where a desire for the data to fit the theory influences the conduct of the review, or anchoring bias, where knowledge or preference around one of two studies influences the review. These are important potential caveats to consider in the use of CCA (although they are not unique to the approach). Some potential ways of overcoming these biases are discussed in Kneale et al. (2015) including the transparent reporting of how the logic model was developed, the depiction of areas of uncertainty a priori, and the inclusion of theorised pathways that were not tested (evidence gaps) in the final logic model based on the review findings. Additionally, a clear articulation of why and how the logic model was updated appears necessary in order to maximise transparency in the review process. In some ways, the expectation of flexibility and iteration between the model and data may help to partially overcome the influence of reviewer bias towards one or two studies, given that this avoids adopting a deterministic approach to systematic reviewing which seeks to narrowly confirm theory, towards a much broader set of questions around how interventions work and which elements can be generalised across settings. As White (2018) usefully reminds his readers, theory should fit the data but data should not be made to fit the theory, and a degree of iteration between model and data is expected, and useful in overcoming some forms of bias.

4.4 Process-Based and Systems-Based Thinking in Logic Models

Recognising that an intervention is complex, and cannot be understood as a single monolithic “whole intervention” is at the basis of systems level theory (Clark, 2013). This also forms the basis of causal chain analysis, as we aim to provide more of a mechanistic account of how interventions effect change, theorising about the complex relationships that may be involved and their interactions with contexts and wider systems.

Most, if not all, social interventions in the field of International Development can be viewed as "systems", which are likely to be "complex", and nested within systems of similar or greater complexity, with interactions taking place between contextual and intervention systems. Systems-based approaches to systematic reviewing involve understanding the ways in which intervention processes and outcomes drive change in a wider system of influence (Rutter et al., 2017). Failure to account for the system of influence in which interventions take place may lead to an incomplete picture. However, in addition to accounting for broad systems of influence, for the purposes of causal chain analyses of international development reviews, all logic models should also be process-based, and articulating the causal relationships between intervention components, and different mediating and target outcomes. Logic models provide a useful starting point as this initial theorising encourages systematic reviewers to consider of all aspects of complexity, in terms of the intervention and the way in which it is nested within the broader system, and how this relationship may differ across contexts (Higgins et al., in press).

Complex social interventions, by their nature, draw upon systems theory for their identification, given that a key characteristic of a complex causal relationship is the dynamic interaction between interventions and broader systems of influence. Building up an 'isolated description' of an intervention's causal chain (Illari and Russo, 2014, Craver, 2001) may be a first step in developing a logic model, but a model that explains the pathway between intervention and outcome (process-based) and considers how the intervention system is nested within a wider system (system-based), is ultimately more useful for reviews of complex social interventions in International Development.

Theorising and synthesising evidence on the way in which broader systems influence the implementation and effectiveness of interventions is also essential in appraising the generalisability of the evidence produced. An initial logic model usually begins through providing a depiction of a causal chain that is broadly generalisable across settings, including some interactions between the intervention and broader system, although with some aspects perhaps being represented as being sensitive to contextual factors and expected to vary across settings/populations. Many of the synthesis techniques presented here (sections 6-8) set out to explore the basis for the broad assumption that the causal chain is generalisable across settings (except where explicitly stated within the logic model), and the evidence uncovered within the systematic review may challenge this understanding. It is possible that the evidence uncovered during the review suggests that an intervention 'works' in such a different way between contexts that separate logic models are needed to represent fundamental differences in the nature of the causal chain across settings. To our knowledge, there are few examples of this sort of cleaving of programme theory occurring on the basis of new evidence uncovered within existing reviews, although this could represent a further advancement in the use of programme theories within systematic reviews in the future.

Section 5

Evidence of Causality in Systematic Reviews Employing Causal Chain Analyses

Section 5 Summary: Evidence of Causality in Systematic Reviews Employing Causal Chain Analyses

- Provided that the underlying assumptions are upheld, randomised controlled trials (RCTs) provide compelling evidence of causation. Systematic reviews drawing on RCT evidence hold potential for providing evidence of causal relationships, but limitations in the type and scope of causal account mean we are unable to provide an explanation as to why or how a change in outcome is achieved. This impedes the 'portability' of the evidence to other situations.
- In order to undertake Causal Chain Analysis in systematic reviews of development interventions, it is necessary to draw upon a plurality of evidence. However, it is also useful to consider how to assessments of whether relationships are causal (or not) are identified.
- Some of the more common approaches for identifying/evaluating relationships as causal are introduced and provide a basis for recognising relationships as causal. Some of the tools cover broader domains than just identifying relationships as causal.
- Tools and frameworks for understanding causal relationships on the basis of qualitative research appear particularly scarce, although some of the approaches that could be employed are summarised, drawing heavily on the work of others before us.

5.1 Causality in systematic reviews of randomised controlled trials

Our epistemological stance with regards to causality tends to reflect both the methods employed in studies included in the review and the way in which this evidence is synthesised. Randomised controlled trials (RCTs) have been regarded as a gold standard in establishing causal relationships (Cartwright, 2007a), and systematic reviews involving meta-analysis of RCTs were placed at the peak of the evidence hierarchy in evidence-based medicine (although such hierarchies can be problematic for social interventions (Petticrew and Roberts, 2003)). RCTs have been described by Cartwright as a deductive approach to establishing causality, given that if the underlying assumptions are met, a positive result

implies causality and clinches the conclusion, rather than merely vouches for it (Cartwright, 2007b, Cartwright, 2007a). She distinguishes between evidence that 'clinches' a conclusion of causality, as in the case of RCTs, and evidence that 'vouches' for a causal relationship, where it 'speaks' of a causal relationship, but where other evidence is needed to secure the conclusion of causality (Cartwright, 2018, p6).

However, the processes undertaken within RCTs are such that they narrow the scope of their application, both in terms of the types of social problems that can be studied, as well as the generalisability of the evidence (Cartwright, 2007b, Cartwright, 2007a, Cartwright, 2010). In addition, there are several ways in which the assumptions of an RCT can be violated, for example breaches in the random assignment to treatment and control groups, which increase bias. Even in well-conducted RCTs, simulation studies suggest chance may still play a role in the determination of the outcome (Clarke and Halsey, 2001); this also extends to the meta-analysis of RCT studies (Clarke and Halsey, 2014). There are also innumerable situations and reasons that arise where conducting an RCT to evaluate the effectiveness of an intervention is unfeasible, inappropriate, or unethical. Where these situations arise, other study designs may be employed that are based on non-randomised experimental designs, although a study employing these designs is arguably less likely to result in evidence that 'clinches' a result with the same certainty as an RCT study design.

Systematic reviews can be useful tools in helping to overcome some of these limitations. Firstly, tools exist to aid systematic reviewers to assess the underlying assumptions of RCTs and other study designs and to assess the risk of bias within a study (Higgins et al., 2011). Where bias is identified, further sensitivity analyses can be employed to explore possible impact on the results. Secondly, with regards to narrow generalisability, for some systematic reviewers, the very act of combining trial effect sizes, which sometimes originate from very different contexts, provides an assurance that the pooled result is 'generalizable'. For example, Donaldson (2001) explains that through synthesising 'different participants in different situations and using different research procedures, one is able to get a better estimate of the robustness or the external validity of a given finding or effect' (p451). Meanwhile, the meta-analysis of effect sizes from non-randomised designs not only broadens the scope of the evidence and its generalisability (as above), but to some extent also strengthens the robustness of the evidence (certainly its potential as a 'voucher') through drawing upon other causal accounts around, for example, regularity accounts in interpreting the evidence.

While systematic reviews of RCTs, particularly those that employ meta-analyses, may hold potential for establishing causal inference, without employing causal chain analysis, we may be less certain why, or how, outcomes are achieved. This narrow scope also limits its potential generalisability to situations beyond those included in the meta-analysis. In the case of non-randomised designs, where the identification of a causal relationship is more challenging regardless, causal chain analyses that seek to understand how an effect was produced also serve to bolster the causal inferences made.

5.2 Identifying causal relationships in systematic reviews that include diverse study designs

Systematic reviews that are reliant on isolated descriptions of interventions will inevitably produce isolated accounts of causal relationships that are potentially much weaker. Such isolated accounts implicitly limit the generalisability of the findings, given that trial mechanisms, particularly for complex international development interventions are, at least partly, context-dependent, and isolated descriptions provide only a partial understanding of these mechanisms. Systematic reviews that have attempted synthesise evidence exclusively from RCTs (or similar study designs that aim to establish narrow "clincher" claims), for complex intervention models, have justifiably been met with criticism for a narrow scope (Petticrew et al., 2016, Cornish, 2015). However, working with more diverse data can introduce challenges in thinking about causality. Criteria or principles for theorising when a relationship is causal can be particularly useful, particularly for reviewers working with more diverse data, and some of these are described below. While useful in considering the properties of causal relationships from evidence that does not provide a 'clincher', many of the approaches included in Box 1 also address questions around the quality of the studies (focussed on internal validity), and not all of the criteria included in some may be suitable for assessing evidence from development interventions.

Box 1: Tools, Checklists and Approaches for Identifying and Evaluating Causal Relationships

GRADE criteria (Guyatt et al., 2008): Although the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) criteria is used in broader terms than causality alone, some elements are especially pertinent to evaluating causal relationships. These include: (i) the consistency of the evidence (whether there is heterogeneity and how much this can be explained); whether a dose-response relationship was observed; whether adjustment for potential confounders occurred; (ii) the size of the effect and the precision of the estimates; (iii) the quality of the evidence and whether the methodological assumptions are upheld within studies; and (iv) whether the findings are generalisable. GRADE is the recommended tool for Cochrane to rate the quality of evidence for systematic reviews of intervention studies (usually RCT studies, although not directly for causality) and the extent to which GRADE should be considered applicable as a tool for evaluating causality has been critiqued elsewhere (Norris and Bero, 2016).

GRADE-CERQUAL (see for example (Lewin et al., 2018)): The GRADE-CERQual ('Confidence in the Evidence from Reviews of Qualitative research') approach provides guidance for assessing how much confidence to place in findings from systematic reviews of qualitative research. These reviews may, or may not, directly address questions of whether and how interventions lead to a change in participant outcomes), although the use of qualitative research to inform decision-making is explicitly acknowledged within the CERQUAL approach. CERQUAL encourages reviewers to

consider: (1) methodological limitations in individual studies, (2) coherence, (3) adequacy of data, and (4) relevance. This tool is being used in Cochrane Qualitative Reviews.

Rogers' strategies (Funnell and Rogers, 2011, Rogers, 2014): Causal relationships are evaluated through three strategies: (i) estimating the counterfactual (i.e. what would have happened in the absence of the intervention, compared to the observed situation); (ii) checking the consistency of evidence for the causal relationships made explicit in the logic model; (iii) ruling out alternative explanations, through a logical, evidence-based process. Some of the strategies for addressing the second of these explore whether intermediate outcomes were also achieved, checking the timing of impacts, undertaking process tracing (e.g. in the case of systematic reviews this could be through undertaking synthesis of process evaluation studies), and checking for dose-response relationships.

Howick criteria (Howick et al., 2009): [Drawn from epidemiology] Causal relationships are evaluated through examining: Size of effect not attributable to plausible confounding; appropriate temporal and spatial proximity (is the interval between intervention and change in outcome consistent with the purported mechanism); dose-responsiveness; plausible mechanism; coherence; replicability (are the parameters of the study comparable); similarity (is the "same" causal relationship being assessed).

Bradford Hill Criteria (Hill, 1965): [Drawn from epidemiology] Causal relationships are evaluated through examining: strength of relationship; consistency (has the same effect been observed multiple times across different settings); specificity (whether the effect is combined in a subset of observations); temporality (does change in the outcome occur after the introduction of the intervention); biological gradient (dose-response relationship); plausibility; coherence (is the effect supported by general theory).

Other criteria are also used across the literature, which also generally involve assessing the strength, plausibility and consistency of causal relationships (Weed, 2000). Some of the frameworks above are based on epidemiological relationships, although many of the individual criteria are relevant to identifying causal relationships in other disciplines. However, using criteria such as described above is not necessarily helpful in identifying some of the more complex causal relationships of the type described earlier in section 3, such as INUS relationships for example. Similarly, many of the approaches above are more suitable for evaluating quantitative evidence and/or the quality of the conduct of the study; for example although GRADE-CERQUAL is named above in relation to qualitative evidence, its purpose is in assessing how much confidence to place in findings from systematic reviews of qualitative research, an aim that could pertain to studies that are descriptive in nature and those that aim to provide causal explanation from qualitative studies. Unfortunately, frameworks for helping to identify or establish causal relationships from qualitative data in systematic reviews are comparatively underdeveloped (Rychetnik et al., 2002); this may be a reflection of the differing epistemologies of qualitative and quantitative research, or may be due to the contribution of qualitative research to causal investigation being historically overlooked (Maxwell, 2004a).

While no single framework for establishing causality in qualitative framework appears to exist (to the authors' knowledge), and such a framework may in itself be unsuitable, Maxwell (2004a, p8) examines the specific properties that qualitative research can bring to causal explanation and in particular the identification of 'causality among particular cases, the importance of context as integral to causal processes and the role of meaning and interpretive understanding in causal explanation'. He describes a 'process-orientated' approach to using qualitative research in causal investigations, aligned with the mechanistic causal accounts described earlier in section 2, as maximising the key strengths of qualitative research for causal attribution. Later, (Maxwell, 2004b), explored how different qualitative methods provide different forms of evidence for causal explanation. While these cannot be directly transformed into a specific framework or checklist for establishing causality, the summaries below (table 1) based on (Maxwell, 2004b) do provide an outline of some of the ways in which causal relationships can be established in qualitative research and checked for their robustness, which in turn could support reviewers undertaking causal chain analyses.

In addition, realist reviews also explore the issue of identifying causal mechanisms more directly and provide a further way of directly identifying causal relationships. For example in Eddy-Spicer et al. (2016) and their review of school accountability in low and middle-income countries, a specific question on mechanisms was included in their data extraction tool that asked reviewers to consider: 'What are the explicit and/or implicit reasons asserted or implied for the connection or disconnection of programme actions to the outcomes of interest (system delivery, system efficiency, and learning outcomes)?' Realist systematic reviews are considered in Section 8 of this paper.

Table 1: Summaries of qualitative approaches, methods or attributes for causal explanation based on Maxwell (2004b)

Type of causal account being developed	Approach, method or study attribute	Description
Observation and analysis of process (aligned with mechanistic accounts described in Section 2)	**Intensive, long-term involvement**	Studies that provide repeated observations where the researcher(s) have been embedded during the course of the intervention can give a clearer picture of causal processes underway.
	Collection of rich data	Collection of data that provide a descriptive account of what happens during an intervention but also provide data to identify the processes and mechanisms involved.
	Narrative and connecting analysis	These studies offer a processual explanation that elucidates the actual connections between events and the complex interaction of causal processes in a specific context; this may be achieved through decomposing and recomposing whole events into sequentially connected social actions [p256 (Maxwell, 2004b)].
Developing and assessing alternative explanations (can be considered as a way to check the robustness of the causal explanation)	**The modus operandi approach**	Studies that would adopt this approach would explicitly theorise and explore alternative explanations that would compromise the proposed causal relationship, and would actively seek out data that could undermine the proposed relationship as a way to check its validity.
	Searching for discrepant evidence and negative cases	Similar to the modus operandi approach above in checking for the existence of discrepant cases, and assessing the validity of those discrepant cases, is a key way of assessing the overall validity of the causal relationship; in some cases, the distribution of cases that support or negate the proposed relationship may be presented.
	Triangulation	Exploring the extent to which causal explanations are supported by diverse evidence collected from different stakeholders and through different methods
	Member checks	Studies that check their interpretations of causal relationships with intervention participants can be considered as having taken steps to avoid misinterpretation of the data, as well as exploring potential alternative interpretations with participants.

Variance Approaches	**Interventions and Comparisons**	This is aligned with interventionist accounts and regularity accounts presented earlier in section 2. Studies employing comparisons may be useful in qualitative research on the impact of interventions, as is the case for quantitative research, for example in exploring the impact of context and whether the same processes and results are observed across different sites in multi-site interventions.

Section 6

Meta-Analysis and Causal Chain Analysis

Section 6 Summary: Meta-Analysis and Causal Chain Analysis

- Meta-analysis involves the quantitative synthesis of outcome data from interventions. It has historically been viewed as having narrow utility for causal chain analyses, as many examples 'lump' together complex intervention processes and contexts. However, there are many counter-examples of studies that use meta-analysis effectively.
- Configurative approaches to meta-analyses are particularly valuable to causal chain analyses when used appropriately, as they can be used to start to trace the development of outcomes across the causal chain and explore potential modifiers.
- Extensions to meta-analysis, for example model-based meta-analysis, are also discussed which show promise to causal chain analysis. A small number of studies, a lack of sufficiently rich data, and potential issues in the modelling do impede their current utility for causal chain analysis, particularly in exploring whole causal chains. However, these extensions and current methods available can be incorporated into complex and robust narratives of causal inference.

6.1 Traditional Approaches to Meta-Analysis in Exploring Causal Chains

Meta-analysis used in CCA involves the quantitative synthesis of data on the direction, magnitude and precision of the impact of interventions on outcomes (see Borenstein et al., 2011). A principle of causal chain analysis (CCA) is that complex interventions cannot be understood as a single undifferentiated "whole" intervention. However, many examples of meta-analysis tend to model interventions as binary exposures, lumping together all intervention processes, and clumping all outcomes as changes that occur simultaneously. For example, Mekasha and Tarp (Mekasha and Tarp, 2013) undertook a meta-analysis of 68 studies examining the impact of international aid on economic growth, finding a modest positive and significant effect on economic growth. The analysis did not shed light on how aid contributed to economic growth, or what forms of aid might be most effective; but in this case the synthesis method was aligned with the research question posed, which sought to settle a controversy over the direction of effect of development aid and not about the mechanisms of action. While the results of such meta-analysis may produce

'more convincing conclusions' (Duvendack et al., 2012), they are based on asking a narrower set of questions than those posed within causal chain analysis.

Configurative approaches to meta-analysis, namely subgroup analyses and meta-regression, can be useful ways of helping to test simple theories about the way in which a limited range of contextual factors, participant characteristics, or intervention components can moderate the impact of an intervention. Meta-analysts using these configurative techniques are, however, repeatedly cautioned that associations observed through such analyses are observational in nature and offer no basis for assuming causality (for example Petticrew et al., 2011, Thompson and Higgins, 2005). These associations are also subject to many of the same caveats of observational research, most notably confounding (Thompson and Higgins, 2005, Sun et al., 2010), although may also be prone to collinearity, and commonly, given that they are based on study-level characteristics, are subject to ecological fallacy in their interpretation (Kneale et al., under review). Nevertheless, this evidence can and is used effectively to develop, or sometimes furnish, hypotheses of what works for whom and in what circumstances.

To take an example of sub-group analyses, a review of land tenure reform established that tenure recognition boosted the land productivity. However, meta-regression suggested that while the impact was positive in all settings, the productivity in sub-Saharan African settings was substantially lower. Perceived tenure security was hypothesised in the initial logic model to be a key mechanism; increased tenure security would then to lead to greater investment in the land leading to increased land productivity. The authors attributed that the smaller effect in sub-Saharan Africa was partly attributable to 'control' conditions where customary tenure relations in Africa tended to provide more security and hence lower risk than those elsewhere (Lawry et al., 2017). Here, the initial logic model helped to justify and interpret the subgroup analysis; the interpretation was also drawn in part from qualitative evidence synthesis which was undertaken in the review.

Further understanding of causal processes can be developed from synthesising evidence for outcomes represented at different points along the causal chain, including mediating factors, which can incorporated into meta-analysis and meta-regression models (Waddington et al., 2012). For example, reviews have presented meta-analytic findings for outcomes along the causal chain for farmer field schools (Waddington et al., 2014b) and deworming (Welch et al., 2017) and used meta-regression to explain heterogeneity in outcomes across studies due to differing rates of adherence in water and sanitation programming (Waddington and Snilstveit, 2009) and deworming (Welch et al., 2017). In the example of farmer field schools, which was described earlier in Figure 1, separate meta-analyses (including sub-group analyses and meta-regression) were conducted on: (i) knowledge levels of farmers, which was theorised to be a necessary condition to adoption of practices; (ii) pesticide use, with lower levels indicative of adoption of Integrated Pest Management approaches, which was a focus of the intervention and theorised to be a sufficient condition for the final outcomes; and (iii) yields and net revenues, which were the final outcomes included in the causal chain.

Despite their utility for testing simple hypotheses, these techniques may be more limited for modelling some of the more complex relationships that can be represented within logic models; furthermore, the number of studies included within reviews often limited the extent of the possible analyses. Extensions to these configurative approaches have been proposed elsewhere. These include a form of enhanced sub-group analysis, undertaken through first exploring similarities between the location in which the evidence is to be applied and where the evidence has been generated, with the differences then forming the basis of sub-group analysis (Kneale et al., 2018). Similarly, using the results of meta-analyses within a mixed-methods framework has also been shown to be effective in uncovering elements of complexity in causal relationships (Thomas et al., 2014), with reviews also being undertaken that model (theory-based) complex combinations of covariates directly within meta-analysis models (Harris et al., in press, Harris et al., 2015). Further, (more established) developments in the conduct of meta-analysis are also becoming useful for the exploring causal chains, and the potential of some of these are described below.

6.2 Extensions to Traditional Configurative Meta-Analysis and their Utility in Exploring Causal Chains

Network meta-analysis allows an analyst to build a network of direct and indirect comparisons between interventions and can be used to test comparative effectiveness of different hypothesised causal chains. An example in literature is a comparison of different approaches to mass deworming interventions and their impact on the developmental health and well-being of children in low-income and middle-income countries (Welch et al., 2017). Here, the authors developed a logic model a priori, which included complex virtuous cycle effects that were expected to operate, and the tested comparative effectiveness of different combinations intervention components in supporting this logic model (e.g. standard pharmacological intervention plus nutritional supplements compared to usual care). In this case, the intervention model was deemed to be ineffective regardless of intervention components, and the use of network meta-analysis provided evidence that 'overall, our analyses do not support causal pathway assumptions about influence of mass deworming on child health and school performance' (Welch et al., 2017, p e41). Despite their potential promise, some of the underlying assumptions of network meta-analysis may be difficult to substantiate and may require additional considerations, particularly for analyses that attempt to include evidence from mixed study designs (although the example above did include evidence from a plurality of study designs, information from quasi-experimental studies was only used in sensitivity analyses).

Other extensions to meta-analysis can also help to mirror some of the complexities in hypothesised causal chains. For example, multilevel meta-analyses allow for modelling of effect sizes while explicitly recognising that these may be organised hierarchically and not

entirely independent of each other (e.g. effect sizes may be nested within sites; sites may be nested within studies; studies may be nested within journals etc.) and allows for the addition of multiple nested effects to be modelled. This is aligned with the systems thinking described earlier in section 4.4. Multivariate meta-analyses are another extension which test intervention effects on outcomes simultaneously, recognising statistical dependence between outcomes from the same study. This approach can be viewed as being aligned with causal chains that describe multiple simultaneous causal strands, and recognise that interventions may need to optimise several causal pathways (Rogers, 2008). Many meta-analytic techniques and their extensions could be enhanced by the use of individual level data for meta-analysis (as opposed to aggregate study-level data). Individual Participant Data (IPD) meta-analysis involves the application of meta-analysis methods to participant-level data and allows more flexible, complex statistical analysis of study data and can enhance the range of causal chain analyses possible (see (Riley et al., 2010)). However, use of IPD meta-analyses remains scarce in the literature reflecting the paucity of IPD from interventions that can be made available to meta-analysts (e.g. through data repositories or shared more directly by triallists); there may also be additional challenges in the analysis of IPD that require specialist skills not available to review teams. Consequently, examples of studies that employ IPD meta-analyses in the field of international development are relatively rare, being confined to observational studies and/or studies focussed on health improvement (for example van Eijk et al., 2013).

Perhaps one of the most direct ways of exploring strands or whole chains using quantitative synthesis is to implement 'model-based meta-analysis' (Becker, 2001, Becker, 2009). As Becker outlines, unlike some of the more traditional approaches to meta-analysis described above, model-based meta-analysis explores whether A leads to B *and* B leads to C (Becker, 2009, p379). Model-based meta-analysis allows for the examination of partial relations, mediating effects, and indirect effects, which are often represented within logic models but rarely modelled in meta-analysis. This form of analysis allows for construction of complex models, similar to structural equation models used in primary literature, and is based on the synthesis of correlation matrices. The results of model-based meta-analyses have been shown to provide a better representation of the social world than using conventional meta-analysis alone. For example, Whitehead and Becker explored the impact of father's involvement in children's upbringing after divorce and uncovered indirect effects that were not detected using conventional meta-analysis, but were supported by theory (Whiteside and Becker, 2000). Becker presents a worked example of the stages involved (Becker, 2009), which are more intensive and require more extensive data than for traditional meta-analyses. Furthermore, few examples exist where such model-based meta-analyses have been conducted on other types of data (e.g. categorical data), although conventional structural equation models on primary data have been generalised to accommodate different data types (Skrondal and Rabe- Hesketh, 2005). In the absence of either IPD data, or sufficiently rich data to support model-based meta-analysis, and potential issues in the flexibility to accommodate different forms of data, systematic reviewers may need to rely on more conventional forms of meta-analysis described above.

These may not provide a causal clinch for the entire causal chain (Cartwright, 2007b), but alongside other forms of synthesis described below, can be incorporated within complex and robust narratives of causal inference (see Krieger and Davey Smith, 2016 for further discussion on incorporating diverse data for causal inference).

Section 7

Alternative Methods and Approaches for Synthesising Data on Causal Chains

Section 7 Summary: Alternative Methods and Approaches for Synthesising Data on Causal Chains

- Framework synthesis is presented as a method for organising or configuring diverse data, including quantitative and qualitative data, to explore causal chains.
- Qualitative Comparative Analysis (QCA) is presented as a method of identifying necessary and sufficient conditions, as well as INUS conditions (see section 3 and glossary) that trigger a successful intervention outcome.
- Theory-based systematic reviews (TBSR) are also examined, with the overlap between TBSR and causal chain analysis (CCA) acknowledged, although a distinction is made with TBSR describing the use of theory more broadly within a review, and CCA more explicitly focussed on the synthesis of evidence for causal inference.

Many different forms of synthesis can aid as part of CCA and other sources provide a detailed account of these (Gough et al., 2017, Snilstveit et al., 2012, Waddington et al., 2012). We describe two synthesis methods/approaches to conducting a systematic review below – QCA and its capacity to identify multiple conjunctural causation, Framework Synthesis to amalgamate different types of data – before exploring realist synthesis in Section 8. We also include an examination of Theory-based systematic reviews (TBSR) as an approach, which shares many of the same principles as CCA, and examples of reviews undertaking TBSR are included to provide examples and inspiration to reviewers considering undertaking CCA.

7.1 Using Framework Synthesis to Organise Different Types of Evidence

Framework synthesis mirrors techniques originally used for analysing large volumes of primary qualitative data (Ritchie and Spencer, 2002), but within systematic reviews of intervention studies have been used as a technique for amalgamating diverse data from quantitative and qualitative studies and for studying complex interventions (Brunton et al., 2015b). Framework synthesis involves five key analytical stages including (i) of familiarisation with the data; (ii) theme identification (creation of a framework for configuration); (iii) indexing of data according to a framework (applying the framework to the data); (iv) charting (rearranging the data according to the framework (and possibly modifying the framework)); and (v) mapping and interpretation of the data. "Best fit" framework analysis involves a deductive phase, where data are synthesised according to the framework, and inductive phases, where evidence that doesn't fit into the framework is also considered (Booth and Carroll, 2015a).

The causal claims resulting from techniques like framework synthesis have been aligned with hypothetico-deductive reasoning (Cartwright, 2007b, Illari and Russo, 2014), where the aim is to uncover enough, sufficiently varied, and novel evidence to substantiate the hypothesis if it were true (Cartwright, 2007b) (in this case that the intervention is in/effective). Cartwright deems hypothetico-deductive approaches to be a more realistic strategy than looking for a single study or sub-set of studies that can provide a casual clinch (Cartwright, 2007b). For CCA it presents a more holistic option in marshalling different forms of evidence to populate different causal strands and through the inclusion of diverse data, framework synthesis can theoretically be used to provide evidence across longer causal chains.

Framework synthesis is a new, but rapidly expanding synthesis method (Booth and Carroll, 2015a, Brunton et al., 2015b). An example includes Brunton and colleagues use of framework synthesis to understand the processes of community engagement and to identify intervention components that support more extensive community engagement through a synthesis of process evaluation studies (Brunton et al., 2015a). A framework, developed from a previous review of community engagement (O'Mara-Eves et al., 2013), was applied to understand community engagement processes, and modified during the course of the review to accommodate new evidence that emerged. Framework synthesis was a particularly suitable method, given the highly variable methods of data collection and analysis that takes place within process evaluation studies. Arguably, this example was focussed on analysing causal chains occurring within interventions – so how components of the intervention led to intervention outputs – and less on how these led to improvements in health status (the outcome of interest).

An alternative example comes from a review of the link between the recent pandemic of Zika virus (a mosquito-borne virus) and congenital brain abnormalities or Guillain-Barré

syndrome (a nervous disorder) (Krauer et al., 2017). This review started with the development of a framework specifically for assessing causal relationships between Zika and adverse child outcomes and nervous disorders. A systematic review was then conducted to assess the validity of the framework that synthesised evidence 'studies of any design and in any language that directly addressed any research question in the causality framework' (Krauer et al., 2017, p5/27), including case reports and case series. Although the approach not explicitly described as 'framework synthesis' by the authors, the description provided appeared to encompass several stages described above. Through developing and testing a framework using hypothetico-deductive means, and evaluation by an expert panel, the authors concluded that Zika virus was indeed a cause of congenital abnormalities and a trigger of Guillain-Barré syndrome. A further example of framework synthesis of development interventions can be found in a review conducted by Hossain et al. (2017) which explored effective approaches for reducing exposure of urban populations to disaster risks.

Framework synthesis is an attractive method for causal chain analysis as it accommodates the synthesis of different types of evidence that may reflect different strands of the causal chain. In addition, when focussed upon similar strands of the causal chain, it upholds other principles in causal attribution, principally triangulation (see section 5). It is also closely related to other techniques, particularly the use of logic models (Harden et al., 2017). However, given that it remains a relatively nascent method, its utility is still being realised, although appears conceptually sound, and its principles reflect the reality of the diverse evidence sources needed to understand long and complex causal chains. Nevertheless, some caveats do apply, particularly around the need to develop standards for practice for the conduct of framework synthesis.

7.2 Capturing Complexity and Providing Regularity Accounts of Causal Relationships through Qualitative Comparative Analysis (QCA)

QCA is increasingly employed as a solution to the challenge of analysing data containing a small number of cases, each with an extensive array of conditions that may trigger a given outcome (Ragin, 2008). This "small N-many variables" challenge is similar to that often faced by systematic reviewers, and Thomas and colleagues provide one of the first examples where QCA was utilised within a systematic review to understand configurations of intervention components that were aligned with "successful" interventions (Thomas et al., 2014). QCA is being used within systematic reviews both to further understand the results of meta-analyses (for example Brunton et al., 2015a), to develop theories to test within meta-analyses (for example Harris et al., 2015), and occasionally as a synthesis method in its own right, although the latter is not encouraged here. QCA allows us to test causal conditions using a regularity account of causality, albeit with mechanistic interpretation. Despite the synthesis ultimately involving numeric data, it is markedly

different from the logic of other forms of quantitative synthesis, with relationships assumed to be asymmetrical, as opposed to the symmetry assumed in statistical relationships (Ragin, 2008). This asymmetry means that, for example, if we find a particular intervention component triggers a successful outcome in a QCA model, the absence of this component does not signify an unsuccessful outcome. QCA has its basis in set-theoretic logic where the focus is on sets of conditions (e.g. intervention components or contextual factors) as entities, rather than the individual constituent components. QCA analyses allow for the consideration two aspects of set relationships, necessity and sufficiency (described earlier) and, building from these, can be used to investigate other complex relationships including multiple conjunctural causation and INUS relationships. In simplified terms, undertaking QCA involves (i) devising rules for operationalising different forms of data into values of 0 or 1 (crisp-set QCA) or between 0 and 1 (fuzzy-set QCA); (ii) creating a "truth table" revealing how different combinations of antecedent condition sets (analogous to variables) overlap with outcome sets; and (iii) using Boolean algebra to reduce multiple configurations of conditions that appear from truth tables to trigger outcomes down to their instrumental parts, to form more parsimonious solutions.

Systematic reviews using QCA as a synthesis method are starting to appear in the International Development literature, with Langer and colleagues applying QCA to understand the critical features of interventions aimed at supporting women's participation in the labour market (Langer et al., 2018). This synthesis was conducted alongside a meta-analysis, and having tested multiple iterations of QCA model, they identified seven conditions that were *necessary* to feature in successful interventions. In contrast, in a review of adult weight management interventions, Sutcliffe and colleagues identified distinct combinations of factors (causal pathways) that were *sufficient* for generating a successful outcome (Sutcliffe et al., 2016); identifying such sufficient relationships is usually the more common purpose and outcome of QCA. When used in combination with other synthesis methods (usually meta-analysis), QCA emerges as a powerful technique of understanding how the organisation of intervention components can cause changes in outcomes.

An example of QCA used in international development

Ton et al. (2017) undertook a review of the evidence on contract farming, which involves establishing a relationship between a farmer and a firm before production begins. It is said to be a popular model with government and donors as was theorised to lead to improved livelihoods for smallholders, and ultimately to economic growth and poverty alleviation. Through quantitative synthesis, the review explored a number of outcomes along the causal chain, although because of methodological and conceptual heterogeneity, ultimately meta-analysed data on increases in income. A series of moderator analyses were undertaken to explore the impact of different factors, which provided a basis for selecting conditions of interest for the QCA. While the moderator analyses provided some indication as to the independent impact of different intervention factors on the effect size, the interaction and combination of these factors and their role in triggering larger effects was examined through QCA. The QCA examined the impact of different configurations of intervention components, for example whether interventions provided seeds or provided credit in cash to farmers, in explaining whether the intervention was highly effective in increasing farmers' incomes. The results were examined by type of crop/produce and a 'price premium' (offering higher-than-local prices) was consistently part of configurations of highly effective studies for annual and perennial crops; for animal husbandry, a package of 'inputs plus credit' was observed to be a condition of highly effective interventions.

7.3 Theory-Based Systematic Reviews

Theory based systematic review (Snilstveit, 2012, Waddington et al., 2014a, Waddington et al., 2012, White, 2009, White, 2018), combines programme theory and mixed methods presentation of evidence along the causal chain in order to explore heterogeneity in findings by context, programme design and implementation. There have been a number of calls for the incorporation of programme theory into systematic reviews over the years (for example, (Pawson, 2002, Davies, 2006, Van der Knaap et al., 2008, Waddington and Snilstveit, 2009, Anderson et al., 2011, Kneale et al., 2015, Maden et al., 2017, Snilstveit, 2012)), as well as calls for multi-disciplinary working (for example (Snilstveit, 2012, Thomas et al., 2004, Oliver et al., 2017, Greenhalgh et al., 2008)). The importance of using theory to develop relevant review questions, structure evidence collection, and present findings is well-recognised (Table 1). These reviews use logic models, theories of change and occasionally middle-range mechanisms or high-level theory.

These reviews have been able to provide answers to some pressing development questions for policymakers and implementers – for example, Table 2 below provides example requirements and findings based on the results of different theory-based systematic reviews on agricultural interventions to improve socioeconomic outcomes. The

examples in Table 2 show how theory based-systematic reviews can help decision-makers and practitioners to identify reasons for successful implementation and participation drawing on evidence participant or implementer views, the effectiveness of targeting, unintended or adverse outcomes for vulnerable groups, or questions about cost-effectiveness. Further examples of reviews that are identified as theory-based systematic reviews are described in Table 3; this shows the application of TBSR across a range of areas relevant to international development.

Table 2: Examples of Hypothesised Decision-Making Requirements and Results from Theory-Based Systematic Reviews (TBSR) for agricultural interventions

Decision-making requirement	**Example Theory-Based Systematic Review evidence**
Using TBSR to understand differential intervention impacts **Example intervention: certification to improve socioeconomic outcomes**	Certification schemes, like Fairtrade, are effective in raising prices and income from agriculture but do not usually improve household income and wages (Oya et al., 2017). Costs of implementing standards can prevent poor farmers joining the schemes, and training is often not oriented to the needs of smallholders and workers.
Using TBSR to hypothesise negative impacts of interventions **Example intervention: Land Reform to improve socioeconomic outcomes**	In addition to some positive impacts, land reform may also have negative consequences, such as conflict, displacement, or reduced property rights for women, as the qualitative evidence in this review indicated (Lawry et al., 2017).
Comparing TBSR to identify the most appropriate approach to engagement and scalability **Example intervention: Agricultural training for farmers to improve**	Top-down agricultural extension does not appear to be effective in improving harvests for African smallholders (Stewart et al., 2015). On the other hand, farmer field schools (FFS), a bottom-up learning approach, improve outcomes along the causal chain (knowledge, adoption, yields, income). But evidence suggests that these programmes do not work at scale due to problems in recruiting, training and back-stopping FFS facilitators (Waddington et al., 2014b).

socioeconomic outcomes	

Any narrow distinction between causal chain analysis and a theory-based systematic review is generally imprecise and may be too abstruse as to serve any real purpose. However, if we consider causal chain analysis to refer to the conceptualisation of the way in which the intervention works, and the choice of appropriate (and likely mixed-method) synthesis approach to understanding part of or the whole intervention, then a theory-based systematic review potentially describes a more encompassing endeavour. Theory-based systematic review describes more explicitly and comprehensively the way in which the conceptual framework developed to represent the intervention is used to design all stages of the review (White, 2018). Here, we have used causal chain analysis more narrowly to refer to the synthesis of evidence and the interplay between the underpinning logic model and synthesis; this usually involves adopting a fine-grained approach to understanding mechanisms and requires understanding interventions as chains of inputs, actions and outcomes (although not always in a linear fashion). Conversely, in addition to being used as a more encompassing term to describe the way in which theory can and should shape the whole systematic review process, theory-based systematic reviews also more explicitly hold an ambition to understand whole intervention systems. As a consequence, a theory-based systematic review may organise evidence in a matrix in order to develop questions for and summarise evidence on whole intervention chains (see (White, 2018) for further information and guidance).

Table 3: Examples of International Development Reviews that Draw on Theory Systematically

Topic or sector	Author	Review title	Type of programme theory used in synthesis stage	Evidence included (synthesis approach)
Agriculture	Lawry et al. (2017)	The impact of land property rights interventions on investment and agricultural productivity	Theory of change	20 quantitative causal studies (meta-analysis) 9 qualitative studies (views)
	Oya et al. (2017)	Effectiveness of agricultural certification schemes for improving socio-economic outcomes	Theory of change	43 quantitative causal studies (meta-analysis) 136 qualitative studies (thematic synthesis)
	Stewart et al. (2015)	Effects of training, innovation and new technology on African smallholder farmers' economic outcomes and food security	Logic model	19 quantitative causal studies (meta-analysis)
	Ton et al. (2017)	The effectiveness of contract farming for raising income of smallholders	Theory of change	22 quantitative causal studies (meta-analysis) 26 qualitative comparative analysis (QCA)
	Waddington et al. (2014a)	Farmer field schools for improving farming practices and farmer outcomes	Theory of change	93 quantitative causal studies (meta-analysis) 20 qualitative studies (thematic synthesis) 337 project documents (portfolio review)
Disability	Iemmi et al. (2016)	Community-based rehabilitation for people with disabilities	Logic model	15 quantitative causal studies (meta-analysis, narrative synthesis)

Topic or sector	Author	Review title	Type of programme theory used in synthesis stage	Evidence included (synthesis approach)
	Tripney et al. (2015)	Interventions to improve the labour market situation of adults with physical and/or sensory disabilities	Logic model	14 quantitative causal studies (meta-analysis, narrative synthesis)
Finance and economy	Brody et al. (2017)	Can economic self-help group programs improve women's empowerment?	Theory of change	13 quantitative causal studies (meta-analysis) 11 qualitative studies (participant views)
Public health	Berg and Denison (2012)	Interventions to reduce the prevalence of female genital mutilation/cutting in African countries	Tabular theory of change with mechanisms articulated	8 quantitative causal studies (meta-analysis, narrative synthesis) 27 qualitative and qualitative studies (narrative)
	De Buck et al. (2017)	Promoting handwashing and sanitation behaviour change in low- and middle-income countries	Theory of change, behavioural theory	42 quantitative causal studies (meta-analysis, narrative synthesis) 28 qualitative ('best fit framework synthesis')
	Waddington and Snilstveit (2009)	Effectiveness and sustainability of water, sanitation and hygiene interventions	Theory of change, diffusion theory	71 quantitative causal studies (meta-analysis, meta-regression, narrative synthesis)
	Welch et al. (2017)	Deworming and adjuvant interventions for improving the developmental health and well-being of children	Logic model	65 quantitative causal studies (meta-analysis, network meta-analysis)
Nutrition	Kristjansson et al. (2016)	Food supplementation for improving the physical and psychosocial health of socio-economically disadvantaged children aged 3 months to 5 years	Logic model	34 quantitative causal studies (meta-analysis) 61 quantitative and qualitative studies (realist review)

Topic or sector	Author	Review title	Type of programme theory used in synthesis stage	Evidence included (synthesis approach)
Schooling	Carr-Hill et al. (2016)	The effects of school-based decision-making on educational outcomes	Theory of change	26 quantitative causal studies (meta-analysis) 9 qualitative studies (framework synthesis)
	Snilstveit et al. (2016)	The impact of education programmes on learning and school participation	Theory of change	238 quantitative causal studies (meta-analysis) 120 qualitative studies (narrative synthesis of barriers and enablers)
Vocational education	Tripney et al. (2013)	Post-basic technical and vocational education and training (TVET) interventions to improve employability and employment of TVET graduates	Theory of change	26 quantitative causal studies (meta-analysis)

Source: Waddington et al. (2018)

Section 8

How are Realist Approaches Used in Reviews of International Development Interventions?

Section 8 Summary: How are Realist Approaches Used in Reviews of International Development Interventions?

- Many realist systematic reviews involve a synthesis of theory as an initial stage in developing a logic model/theory of change in a more in-depth way than described in Section 4, which is then used as an anchor for the remaining synthesis.
- Realist systematic reviews encourage reviewers to think directly about causality through placing an emphasis on mechanisms, rather than interventions or programmes, as the central unit of analysis and the means of achieving a change in outcomes.
- Realist reviews seek to identify configurations of context-mechanisms-outcomes (CMO) to explain how interventions lead to a change in outcomes. Different types of evidence (e.g. quantitative and qualitative) are synthesised in order to identify these CMO configurations.
- Realist reviews are presented as a separate section as the preceding stages before reaching the synthesis stage is also conducted differently within a realist review (e.g. searching and screening).

Unlike systematic reviews examining the effectiveness of interventions, realist synthesis aims to unpack the complexity of programme theory and understand how the programme can produce particular outcomes. The concept of “generative approach to causation” adapted by realist perspectives implies that various causal mechanisms, rather than “programmes”, are the unit of analysis and the key to generate desired changes (Pawson, 2002). Cognitive or emotional reasoning of different intervention actors and resources available can be seen as a driving force for triggering changes, which vary according to particular circumstances. By identifying causal mechanisms (M) that lead to the desired outcomes (O) and tracing back to relevant conditions (C), it offers an explanatory power that goes beyond answering “what works” question but explaining “why it happened, for whom and under what circumstances” (Pawson, 2002). This “configurational thinking” can

inform policy and practice in the field of international development, where evidence of impact may be inconclusive, through providing insights into the design of interventions that include the "ingredients" necessary for programmes to work (Pawson and Tilley, 1997, Mallett et al., 2012).

Realist synthesis has been conducted more broadly in public policy and health-related fields but less commonly in international development where context is "the primary consideration" (Mallett et al., 2012, p452). As outlined by Pawson (2002), building on similar causal mechanisms operating under different contexts provides insights on how to implement successful interventions. "Realist reviewing" describes different realist approaches to evidence synthesis, each aiming to undercover how programmes lead to (un)expected changes (see Table 4 for further details and examples from international development (Dieleman et al., 2009, Kane et al., 2010, Westhorp et al., 2014, Eddy-Spicer et al., 2016)). Other examples of CMO synthesis being integrated into systematic reviews include reviews of interventions to prevent female genital mutilation/cutting (Berg and Denison, 2012); interventions providing supplementary feeding to improve child health (Kristjansson et al., 2016); and interventions to improve access to microcredit (Vaessen et al., 2014).

Table 4: Examples of Realist Reviews of International Development Interventions

Study	Intervention focus	Initial theoretical framework	Types of evidence included	Measures to assess rigour* and transparency	Process of identifying and configuring C-M-O
Dieleman et al. (2009)	Human resources management (HRM) interventions Seven types of interventions in scope and classified according to the three HRM-intervention levers Research Question: Which Human Resource Management Interventions in LMICs are successful in improving performance under which circumstances and for which	Developed a framework to facilitate understanding of mechanisms which shows that there are variety of relevant mechanisms Included studies that did not report on the underlying assumptions of how the interventions should bring about to change.	All types of study design	Bias in the evaluation studies	'We systematically assessed outcome, context, and mechanisms through which the intervention produced its outcomes.' (Dieleman et al., 2009, p2) Mechanisms were identified if the study authors reported them. The review teased out three mechanisms that were triggered by HRM interventions and brought about change in health workers' performance, although mechanisms were only discussed to a limited extent and even to a lesser extent researched. Mechanisms included increased knowledge and skills, improved motivation and feeling of being obliged to change. Considered theories of behaviour change Limited reporting on the context, implementation, mechanism, underlying assumptions of how the intervention should bring about change.

	groups of health workers?				
Kane et al. (2010)	Use of community health workers (CHW) Research Question: How and when are community health workers effective for in management of childhood illnesses (IMCI)?	Not stated	RCTs	Not reported	Mechanisms were included only when they were either researched or discussed by the authors of the RCTs (Kane et al., 2010, P4). . Iterative and discussion between review teams, a common understanding of C-M-O was arrived Examples: "Interventions involving better positioning of the CHW within communities (e.g.: Selection of the CHWs in consultation with beneficiary communities; the CHWs being members of the beneficiary community, and perceived by them as role models) can improve the CHW's performance when they are able to trigger the following mechanisms: • an anticipation of being valued by the community, • a perception of improvement in social status, and having a valuable social role • a sense of relatedness with and accountability to the beneficiaries"
Westhorp et al. (2014)	Community accountability Research question: Under what	Draft programme theory developed during protocol stage	All types of studies	Trustworthiness of data within reports	Developed programme theory, drafted a hierarchy of outcomes, described mechanisms as: actors whose decision-making has been changed, the reasoning that underlies the changed decision, and outcome of the different decision, and refined the initial programme theory

	circumstances does enhancing community accountability and empowerment improve education outcomes, particularly for the poor?				Example: "In the Philippines Textbook Program, there was a strong incentive for suppliers to get the delivery correct, as any rejected shipments had to be rectified at the publisher's expense (Majeed 2011; p. 10). There is evidence that the imposition of consequences for poor performance led to improved performance "
Eddy-Spicer et al. (2016)	School accountability systems: assessment, monitoring and inspection Research Question: Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised?	Initial rough theory was developed at the scoping exercise stage, consulting with advisory group members	All types of study designs	Rigour and Relevance	Iterative process involving five rounds of data synthesis, the final round consisted of a comparison across all school accountability elements: assessment, monitoring, and inspection. The review team coded all the included studies on C-M-O. Then, they generated descriptive codes in more details after read and reread coding and full-text papers again. They further clarified conditions that facilitated or impeded the outcomes. The final round employed constant comparative methods to consider mechanisms and make inferential claims Example: "High-stakes examinations are more likely to increase efforts by individual teachers on exam preparation and working with lower performing students and produce sustained increases in test results (O) **through the desire for reward (M).** The evidence suggests that this is more likely to be the case when there are (C): teacher-level individual incentives,

					pressures from school leadership and external stakeholders for results, or teachers' recognition that the incentive is of value and merits additional effort.

*Rigour: whether the methods used to generate the relevant data are credible and trustworthy

Identifying and developing the theoretical framework at the onset of the review process is generally a first step in defining scope of the review and identifying generic causal mechanisms before the synthesis of evidence takes place (Snilstveit et al., 2012, Eddy-Spicer et al., 2016, Westhorp et al., 2014, Dieleman et al., 2009). For example, a recent review of school accountability systems developed an initial theoretical framework after conducting a scoping exercise, consulting with experts in the field, utilising knowledge expertise within the review team. Here, five key generic mechanisms were identified explaining how school accountability systems do (or do not) lead to improved service delivery and learning outcomes of students from developing countries. Studies included in the synthesis were then interrogated to identify the connection between contextual information in the local school context and the particular outcomes, guided by the initial theoretical framework. Similarly, in a systematic review of Human Resource Management interventions (described above (Dieleman et al., 2009)), a framework was developed to facilitate understanding of mechanisms of human resource management interventions to improve availability, productivity, responsiveness, and competency of workers' performance in low and middle-income countries.

Quality appraising in systematic reviews aims to evaluate whether the methods employed are appropriate and the findings are reliable (Gough et al., 2017). Whilst realist synthesis considers "rigour", it also recognises quality "an emergent property" (Eddy-Spicer et al., 2016, p22) throughout the process of review (Pawson, 2006). In addition, relevance is considered by the extent to which the findings support or refute the initial theoretical framework (Eddy-Spicer et al., 2016, Greenhalgh, 2014). The process of generating C-M-O configurations, and constructing or refining the theoretical framework, is iterative and interpretive in nature, working between review team members whilst working on data extraction and data synthesis in order to understand and identify C-M-O configurations (Kane et al., 2010). For example, reviewers typically report several rounds of reading and re-reading data, then comparing and contrasting related features of C-M-O configurations across different interventions, before developing a more refined theoretical framework that explains how programmes lead to the change in particular outcomes (Eddy-Spicer et al., 2016, Kane et al., 2010, Westhorp et al., 2014). It also requires review teams to engage with different types of evidence to identify the connection between context, mechanism, and outcomes that would provide essential information for establishing potential inferential claims.

Section 9

Conclusions: Causal Chain Analysis in Systematic Reviews of International Development Interventions

Section 9 Summary: Causal chain analysis, principles for best practice and challenges and strategies

- Causal Chain Analysis (CCA) does not describe only one form of evidence synthesis, but involves the selection of synthesis methods that are able to address different research questions about causal chains and how interventions lead to a change in outcomes. These different synthesis approaches inevitably draw about different causal accounts.
- CCA always involves the development of a logic model prior to undertaking the review, upon which decisions about synthesis are subsequently based. Guidance for developing logic models exist, and logic models benefit from the involvement of stakeholders. Research questions addressed through CCA usually involve drawing upon and synthesising a plurality of different types of evidence.
- No set criteria exist for best practice in the conduct of CCA although we recommend the following as a set of principles: (i) reviewers should be familiar with the underpinning assumptions of CCA; (ii) all CCA involve development of a logic model; (iii) research questions for synthesis should draw on hypothesised causal chains represented in the logic model; (iv) synthesis methods should be selected based on the type of hypothesised relationships that are identified within the logic model; (v) integration of different forms of evidence serve to strength the mechanistic account of how interventions lead to change; (vi) logic models should be updated at the end of the review to reflect the review's findings.
- Some of the limitations of CCA, including the oversimplification of complex causal pathways and the influence of potential defects in the initial logic model, should considered by reviewers undertaking CCA. Some of the ways of overcoming these limitations are discussed above.

Taking a CCA approach enables reviewers to start overcoming some of the critiques that have been levelled at systematic reviews of international development in the past, and particularly the element of "context stripping" of evidence (Cornish, 2015). Understanding interventions as causal chains and examining the mechanisms of action that form the chain links and the optimal organisation of intervention components and contextual and other moderators, as well as factors around implementation and take-up, can be a first step in aiding reviewers to conceptualise the degree to which interventions may generate complex causal relationships. In her wide-ranging critique of systematic reviews of international development interventions, Cornish draws on her own experience of conducting a systematic review, which included only quantitative studies, and calls for '… a broadening of the understanding of "evidence" beyond the prioritisation of systematic reviews and RCTs.

Local case studies of intervention processes in context, theorisations of practice, experimentation with novel intervention processes, perspectives of local people - these are all sources of information that do not contribute to EBP [evidence-based policy-making] as currently defined, but which build valued intellectual resources for informing action' (Cornish, 2015, p273).

While systematic reviews of RCTs may have historically been given priority in decision-making, synthesis of a broad range of types of qualitative evidence has flourished over recent decades (Gough et al., 2017, Thomas and Harden, 2008) with new approaches continually developed (Sutcliffe et al., 2015, Thomas et al., 2014). However, where arguments made by Cornish align with some of the points made in this paper is that analyses of full causal chains are likely to require a plurality of forms of evidence and may need to draw on several causal accounts, in order to evaluate different strands or segments of the causal chain. This may include drawing on qualitative research for causal explanation. No one synthesis method alone is likely to provide a complete causal account of the processes linking intervention inputs, outputs and outcomes; this is in much the same way that philosophers advocate that "evidential pluralism" can strengthen causal hypotheses (Krieger and Davey Smith, 2016, Reiss, 2009, Reiss, 2012). This is similar to some of the ideas advocated within mixed studies/mixed methods reviews, although Causal Chain Analyses might be flexible in focussing on the integration of qualitative and quantitative data to scaffold the same "link" (Pluye and Hong, 2014), as well as on exploring different forms of causal relationship, at different points in the causal chain, and their potential moderators. Similarly, while CCA may share some ambitions with realist reviews, there is scope within CCA for accommodating a number of different synthesis methods including meta-analysis, and bringing together different types of causal reasoning.

A summary of the methods described in this paper and their utility for CCA is provided below in table 5, although as has been stated earlier, other synthesis methods are also appropriate for causal chain analysis including narrative synthesis (Snilstveit et al., 2012), many different forms of qualitative evidence synthesis (Thomas and Harden, 2008, Sutcliffe et al., 2015), as well as different forms of quantitative analysis and meta-analysis not discussed specifically here (Higgins et al., in press). Those included in table 5 and discussed earlier in this report are featured as they either show promise for causal chain analysis but are comparatively rare in systematic reviews of development interventions; or, as in the case of meta-analysis, there is further scope for their repurposing to undertake CCA.

Table 5: (Non-exhaustive) List of synthesis techniques and approaches that could be applied to causal chain analyses of systematic review interventions

Synthesis approach	Benefits for CCA in systematic reviews of international development interventions	Disadvantages for CCA in systematic reviews of international development interventions	Example research question Key text/further reading
Meta-analysis	Can provide compelling evidence for causal attribution provided assumptions of underlying studies and model are met, particularly for the meta-analysis of RCT studies. Can be used to explore different outcomes along a causal chain (usually in independent models), although there are comparatively few examples of effective use of meta-analysis to purposefully explore outcomes at different points in the causal chain.	In the absence of configurative analysis, does not provide any indication of how context, participants or setting influence the size or direction of evidence differentially. Limited to synthesising evidence of more simple theorised relationships. The number of studies available usually precludes analysis that is more detailed.	Example Research question: How effective is intervention x in improving outcome y? See Borenstein et al. (2011) for a detailed overview of different meta-analysis approaches and Higgins et al. (in press) for an overview of utility for complex interventions.
Meta-analysis – subgroup and meta-regression	Can provide compelling evidence for causal attribution provided assumptions of underlying studies and model are met, particularly for the meta-analysis of RCT studies. Can be used to explore different outcomes along a causal chain and can include testing of simple theorised contextual moderators, although as above, there are comparatively few examples of	As above, limited to synthesising evidence of more simple theorised relationships. The number of studies available usually precludes analysis that is more detailed. Complex mechanisms challenging to model/identify, in the absence of a mixed-methods component of the review.	Example Research question: Does the effect of intervention x in improving outcome y differ across setting/participants? See Borenstein et al. (2011) for a detailed overview of different meta-analysis approaches and Higgins et al. (in press) for an overview of utility for complex interventions.

	effective use of meta-analysis to purposefully explore outcomes at different points in the causal chain.		
Meta-analysis extensions – model-based analysis and network meta-analysis	Network meta-analysis – can be used to compare the effectiveness of multiple differing intervention types (or causal pathways). Model-based meta-analysis - can be used to model more complex causal pathways and relationships. Both provide compelling evidence for causal attribution provided assumptions of underlying studies and models are met. In the case of model-based meta-analysis can potentially be used to focus on mediators and attrition or amplification of effects, providing insight into some of the complex mechanisms described in Section 3.	Few examples exist in the international development literature. Undertaking such analyses is dependent on obtaining sufficiently large and rich data, which is challenging. Potential issues in the flexibility of data that can be incorporated in model-based analysis and the types of causal pathway that can be compared in the case of network meta-analysis. Network meta-analysis in particular may give restricted indications on how/why differences are apparent.	Example Research question Network Meta-Analysis: What is the comparative effectiveness of interventions/pathways for outcome y? Example Research question Model-based Meta-Analysis: Does intervention x lead to an improvement in outcome y1 and also then in y2 and y3? See Higgins et al. (in press) for an overview of utility for complex interventions; see Li et al. (2011) for an overview of potential and disadvantages of Network Meta-Analysis; and Becker (2009) for more detailed examination of model-based meta-analysis.
Qualitative Comparative Analysis	A synthesis approach well placed to explore the existence of necessary, sufficient and INUS relationships in data; one of the few methods available for systematic reviewers to examine this complexity with small datasets.	A new technique and few examples exist in the international development literature. Reliant on regularity accounts of causation which may not provide sufficient evidence to clinch a conclusion of causality	Example Research question: Which intervention components and contextual characteristics trigger successful outcomes? See Thomas et al. (2014) for an outline of the use of QCA in systematic reviews
Framework Synthesis	A synthesis approach that can provide a mechanistic account of how interventions work involving	Few examples exist in the international development literature	Example Research question for synthesis: Which intervention causal pathways are confirmed (or refuted)

	the synthesis of diverse data (see section 7.1).	(and beyond). Standards for good practice and conduct unclear.	and which newly emerge when undertaking a framework synthesis of intervention x on outcome y? See Booth and Carroll (2015a)
Realist synthesis	Able to provide evidence for complex and contextually dependent mechanisms described in Section 3. Places analysis of causal pathways at the centre of the review through identifying configurations of context-mechanism-outcomes in the data	Few examples exist in the international development literature (and beyond). While diverse data can be synthesised in the identification of CMO configurations, presentation of results is usually narrative or tabular and quantifiable estimates of mechanism and its impact is not always clear.	Example Research question: Does intervention x work differently across different populations and settings and why do differences arise? See Pawson et al. (2005) for an overview of realist systematic reviews.
Theory-based synthesis	Aligned closely with CCA, although broader in scope than pertaining to the synthesis of evidence on causality, as is the case for CCA.	There are a growing number of examples of this approach, and links with CCA and framework synthesis could be better articulated. Standards for good practice and conduct are unclear.	See White (2018)

9.1 Principles for Best Practice in the Steps Undertaken Within Causal Chain Analyses Included in Reviews of International Development Interventions

No set guidelines exist for the conduct of Casual Chain Analysis (CCA), although guidance does exist for the conduct or reporting of different synthesis approaches (see (Gough et al., 2017) for an overview) as well as principles for developing logic models and theories of change (White, 2018, Kneale et al., 2015). A further paper in the CEDIL series also provides a comprehensive insight into the improvement of model-based conceptual frameworks such as theories of change and logic models (Davies, 2018). The following represent loose principles that could be applied in the conduct of future CCA for International Development systematic reviews.

1. **Familiarity with underpinning assumptions:** CCA describes an approach not a sole method of synthesis. Invoking CCA necessitates an ambition to understand whether interventions work, but also why and how they work. The interventions in scope for CCA are likely to be both complicated and complex, with some mechanisms being partly or entirely context-dependent in their triggers. "Systems-thinking", and viewing interventions as systems nested within larger systems, can be instrumental in establishing some of the relationships that may be moderated by the context in which the intervention takes place.
2. **Development of a logic model to anchor the review:** All CCA are guided by logic models. The steps around the development of logic models were described earlier. Additional elements of good practice include: the development of several iterations and agreement across the review team and its advisors; the representation of potential complex causal relationships that may operate; providing a full representation of the major causal chains and moderators (which also needs to be balanced against keeping a manageable number of permutations of pathways that could potentially be explored (Davies, 2018)); the involvement of intervention stakeholders in the development of the logic model; the representation of potential harms (dark logic (Bonell et al., 2014)); the representation of contextual factors; and the extensive use of the logic model to guide and interpret the synthesis (Kneale et al., 2015, White, 2018). Perhaps one of the most important principles is to secure the involvement of a range of stakeholders in the development of the logic model (see CEDIL paper by Oliver et al on optimal ways of engaging stakeholders) in order to strengthen the salience of the model and its value in subsequent causal chain analysis.
3. **Development of research questions that relate to hypothesised causal relationships and chains:** Research questions should be developed that avoid treating the intervention and/or outcomes as monolithic "wholes"; this does not necessarily equate to avoiding "what works" questions altogether but expands on

these questions to make them specific to particular causal pathways or sets of intervention components.

4. **Justification of synthesis method and study type:** Study types and syntheses methods should be selected that are based on the type of hypothesised relationships that are identified within the logic model, and which address the overall research questions. Reviewers should (be encouraged to) communicate the implications of the selection of different modes of synthesis in terms of the causal accounts that developed, and the type of causal reasoning that might be exercised in interpreting the evidence (and where gaps may lie).
5. **Integration of different forms of evidence using different modes of synthesis:** To better capture longer and more complex causal strands, CCA ideally will involve different forms of evidence and different modes of synthesis to develop a mechanistic account of if and how interventions "work". Where this is not possible, for example, because of limitations in the evidence base for primary studies or because of other constraints, potential gaps and limitations in the CCA should be identified and clearly reported with reference to the logic model.
6. **Updating the logic model to reflect new evidence uncovered during synthesis:** Once a review has identified the underlying causal pathways linking intervention components with different outcomes, this evidence can in many cases be used to update the logic model, either through changing some of the assumptions about how an intervention works or/and through representing the strength of evidence. Willey and colleagues present an effective example where a logic model was updated to reflect the strength of evidence for different causal pathways in a systematic review on the effectiveness of interventions to strengthen national health service delivery on coverage, access, quality and equity in the use of health services in low and lower middle income countries (Willey et al., 2013, p83). This also showed which pathways were not assessed during the review process.

9.2 Challenges and Strategies for Causal Chain Analyses

Some of the challenges facing users of CCA include that no one method of synthesis discussed here is likely to provide a conclusive mechanistic account of how and how much an intervention changes an outcome. Synthesising different data may be one strategy to overcome this limitation (see also White, 2009), and particularly adopting synthesis methods/approaches such as Framework Synthesis that provide ways of integrating these data. Realist reviews are another analytical framework for understanding how context sensitive some combinations of mechanisms and outcomes can be but often omit quantitative synthesis. Strategies such as realist synthesis and framework synthesis are contingent on a rich and varied evidence base, which may not exist for some interventions. The utility of model-based meta-analysis was also explored, and this paper also discussed the possibility of better or more creative deployment of existing (single) synthesis methods, for example the use of covariates reflecting complex conditions directly within meta-

analysis (Harris et al., in press, Kneale et al., 2018). Some of these approaches are also related to the use of QCA, which was identified as a powerful technique in understanding optimal conditions for the organization of intervention components.

Economic synthesis, using sophisticated statistical modelling to derive an intervention's true impact and estimate its cost-effectiveness, and presented in a policy-friendly format, may ostensibly be of greater interest to policy-makers than some of the mechanistic accounts described here (White, 2014). But without an understanding of how the intervention works, such evidence of cost-effectiveness becomes the type of evidence that provides the "clincher" (Cartwright, 2007b), but in such narrow terms that its application elsewhere is challenging. For international development, where contextual factors of importance are diverse and important, "clinchers" become of limited value for future decision-making without understanding the underlying processes. The techniques described in this paper help to establish and enhance the salience of systematic review findings across settings, helping to meet CEDIL's terms of reference around 'systematically and rigorously accumulating, modelling and analysing bodies of evidence in a manner that improves the external validity of findings and identifies where further investigation is most needed' (DFID, 2017).

In this paper we describe CCA as involving the development of a logic model and its use to anchor subsequent analysis, which aims to provide empirical evidence for parts of the causal chain and information about contextual modifiers. This approach, and certainly the ambition of understanding if, how and where interventions work, ostensibly has similarity and overlap with other approaches described here, including theory-based systematic reviews (TBSR) and realist synthesis. We have attempted to make a distinction between CCA and TBSR, describing the latter as distinctive in trying to use theory to guide all stages of producing a systematic review and usually involving an ambition to understand an intervention more holistically, whereas CCA may focus on particular causal strands. However, this distinction should perhaps be regarded as tautological, and both approaches should be viewed as being complementary, with CCA the main approach to analysis within a TBSR. Similarly, the distinction between CCA and realist reviews can also appear, at first, tautological, particularly with the focus on context and mechanisms. Here we would view realist synthesis as an approach that could be taken in undertaking a CCA, but whereas the actual combining of evidence (synthesis) in a realist synthesis is usually narrative in scope (Rycroft-Malone et al., 2012), CCA can refer to a number of different synthesis models, that could be undertaken in the same review. Furthermore, although realist synthesis has increasingly been associated as the default synthesis method for review questions that focus on the way in which interventions interact with context (Kane et al., 2010), as many of the examples in this paper show, other synthesis methods can also unpack these relationships.

The prominent role that a logic model plays within causal chain analysis may theoretically lead reviewers to prioritise seeking evidence confirming their prior assumptions, over an exploration of unintended consequences or to construct new explanations for intervention

effects (even where reviewers have used “dark logic” to theorise negative intervention impacts). Antidotes to this potential confirmation bias can be sought from (i) the involvement of different stakeholders in the development of a logic model who may be able to articulate where generative explanations may be most useful; (ii) explicit representation of ambiguity within certain parts of the causal chain where new theory/explanation is most valuable; and (iii) utilising synthesis methods that support both confirmation of existing assumptions as well as developing new explanations, many of which have been outlined in this paper. Similarly, the logic model is a guiding hypothesis, often based on poor or incomplete descriptions of interventions (Hoffmann et al., 2017), and despite incorporating the elements of good practice described above, may oversimplify (and thus incorrectly specify) a complex systems-based intervention (Rogers, 2008). Here, arguments made by Rogers are useful in recognising that ‘the art of dealing with the complicated and complex real world lies in knowing when to simplify and when, and how, to complicate’ (Rogers, 2008, p30). CCA allows us to theorise the complicated and complex; to hone in on particular parts of the chain (simplify) or to attempt to understand longer strands (complicate) and provides us with the potential to confirm existing theories, or to develop entirely new ways of understanding how interventions effect change.

Section 10

Recommendations for DFID and CEDIL

This paper has presented Causal Chain Analyses as a means of understanding intervention impacts and how they occur. This broad goal aligns closely with CEDIL’s terms of reference in contributing to the advancement of emergent, pioneering and cutting edge methodologies for impact evaluation, in order to maximise the effectiveness of spending on international development (DFID, 2017). Many of the techniques described in the paper are specifically designed to explore the generalisability of the findings, helping to meet CEDIL’s terms of reference around systematically and rigorously accumulating, modelling and analysing bodies of evidence in a manner that improves the external validity of findings and identifies where further investigation is most needed (DFID, 2017).

The flexibility of CCA and the ability to synthesise data that allows for understanding processes of implementation and their links with interventions effectiveness, within a unified causal chain, also aligns with CEDIL’s aim around promoting the use and uptake of evaluation evidence in international development organisations and their partners (both in the UK and internationally) (DFID, 2017). In particular, the way in which stakeholder involvement can be integrated into different stages, and particularly in developing an initial logic model, can allow CEDIL to move more easily “beyond the evidence” and ensure that policy-relevant concerns are integrated into decisions about the focus of a systematic

review and the type of synthesis. Similarly, the capacity to use logic models to theorise about unintended intervention consequences (Bonell et al., 2014) also improves the usefulness of systematic review evidence for identifying where interventions are not appropriate (as well as being aligned with moves in popular science to understand and sometimes celebrate causes of intervention failure (Bansal, 2012)). Despite the apparent benefits, using logic models to understand intervention effects and why they happen is an underutilised strategy in systematic reviewing (Kneale et al., 2015), and as a consequence, there remain many areas in which methods and approaches are underdeveloped. Addressing some of these may be where the work of CEDIL can make a methodological impact, as well as the anticipated substantive impacts in improving the evidence available for decision-makers. Alongside the recommendations below, DFID should also consider investing in the development of further training to enhance the capacity of systematic reviewers to undertake CCA in systematic reviews, particularly in settings where there are known to be existing capacity issues (Oliver et al., 2015).

1. **Undertaking systems-based intervention reviews**: Recognising that complex causal chains within interventions are composed of nested systems within a system, which is itself may be complex, allows for greater conceptualisation of how contextual factors and complex mechanisms may take place within an intervention. However, while steps have been taken towards describing these systems, the tools to analyse and synthesise evidence on a system level are lacking (Rutter et al., 2017). There is an opportunity for CEDIL to support further methodological development in this area, particularly given the arguments made in this paper around how international development interventions fit within this conceptualisations of a complex system. From a quantitative perspective, this may involve greater development of model-based meta-analysis methods (and other alternative approaches to meta-analysis) although a deeper understanding of the causal chain and a more robust causal account, as has been advocated throughout this paper, will follow from methods incorporating a plurality of sources.
2. **Incorporating macro-level theories into reviews**: Development interventions take place in contexts where there are wide differences in political, economic and cultural forces that shape outcomes. Across different disciplines, the emergence and maintenance of these systems are described and explained by different macro-level theories, although these rarely feature within extant causal chain analyses and (related to exploring the utility of system-based reviews above), grounding causal chain analyses in macro-level theory, as well as more micro-level theory, may help CEDIL to understand why and how interventions work.
3. **Greater focus on undertaking reviews of implementation**: Casual chain analysis involves synthesising evidence for part or whole intervention chains, including the way in which different strategies for implementation may lead to outcomes. There are few extant reviews that focus on routes to successful intervention implementation, despite many of the synthesis methods described here being useful for considering these questions. Similarly, different research questions may

also arise when focussing on implementation, for example when trying to scale up an intervention, requiring the analysis of a very different causal chain with very different financial, human, and capital resources needed and mechanisms of operation. A contribution of CCA within a CEDIL framework would be to undertake reviews that give greater weight to understanding evaluating delivery mechanisms.

4. **Supporting the development of methods to assess the transferability of interventions and empirical research exploring the generalisability of evidence:** Many of the existing methods for assessing the generalisability of interventions based on causal chain analysis tend to focus on considerations of feasibility. However, there are a number of instances where it may be possible to run an intervention, although the anticipated impacts may be different because of the characteristics of the setting or participants. New methods are needed in order to support assessments around the likely impact of interventions across different (and potentially unobserved) settings, taking into account the potential impacts of contextual factors on the causal chain. There also needs to be further empirical research into exploring the generalisability of findings. This could be based on further explorations of if/how CCA findings from a subset of studies in a review apply and/or predict the intervention impact in a different subset of studies. Other proposed methods have been described elsewhere (Kneale et al., 2018), although require greater leadership from organisations such as CEDIL to test and adapt these. In the case of CEDIL, this is would be the very type of organisation most likely to benefit from methodological development in the area of generalisability given the complexity of interventions in scope.
5. **Further development of quality assessment tools for causal chain analysis:** Despite the increasing plethora of guiding principles for best practice in developing a logic model (Anderson et al., 2011, Kneale et al., 2015, Maden et al., 2017, Rehfuess et al., 2017, White, 2018), there exists no unified way in which logic models can be quality assessed, either around the steps undertaken in their development, or their suitability or utility in hypothesising or describing intervention effects. Similarly, despite the principles around the steps undertaken in CCA outlined earlier, there is no quality assessment tool for causal chain analysis, or a tool to assess the appropriate use of theory in systematic reviews. This is an area where CEDIL could make a substantial contribution and help strengthen the quality of the evidence synthesis being conducted for international development.

References

ANDERSON, L. M., PETTICREW, M., REHFUESS, E., ARMSTRONG, R., UEFFING, E., BAKER, P., FRANCIS, D. & TUGWELL, P. 2011. Using logic models to capture complexity in systematic reviews. *Research synthesis methods,* 2**,** 33-42.

BANSAL, S. 2012. The power of failure. *The New York Times*, November 28th.

BAXTER, S. K., BLANK, L., WOODS, H. B., PAYNE, N., MELANIE, R. & ELIZABETH, G. 2014. Using logic model methods in systematic review synthesis: describing complex pathways in referral management interventions. *BMC Medical Research Methodology,* 14**,** 62.

BECKER, B. J. 2001. Examining theoretical models through research synthesis: The benefits of model-driven meta-analysis. *Evaluation & the health professions,* 24**,** 190-217.

BECKER, B. J. 2009. Model-based meta-analysis. *In:* COOPER, H., HEDGES, L. V. & VALENTINE, J. C. (eds.) *The handbook of research synthesis and meta-analysis.* New York: Russell Sage Foundation.

BENJAMIN-CHUNG, J., ABEDIN, J., BERGER, D., CLARK, A., FALCAO, L., JIMENEZ, V., KONAGAYA, E., TRAN, D., ARNOLD, B. F. & HUBBARD, A. 2015. The identification and measurement of health-related spillovers in impact evaluations: a systematic review. London: International Initiative for Impact Evaluation (3ie).

BERG, R. C. & DENISON, E. 2012. Interventions to reduce the prevalence of female genital mutilation/cutting in African countries. *Campbell Systematic Reviews,* 8.

BONELL, C., JAMAL, F., MELENDEZ-TORRES, G. J. & CUMMINS, S. 2014. "Dark logic": theorising the harmful consequences of public health interventions. *Journal of epidemiology and community health,* 69**,** 95-98.

BOOTH, A. & CARROLL, C. 2015a. How to build up the actionable knowledge base: the role of 'best fit'framework synthesis for studies of improvement in healthcare. *BMJ Qual Saf,* 24**,** 700-708.

BOOTH, A. & CARROLL, C. 2015b. Systematic searching for theory to inform systematic reviews: is it feasible? Is it desirable? *Health Information & Libraries Journal,* 32**,** 220-235.

BORENSTEIN, M., HEDGES, L. V., HIGGINS, J. P. T. & ROTHSTEIN, H. R. 2011. *Introduction to meta-analysis,* New York, John Wiley & Sons.

BRODY, C., HOOP, T. D., VOJTKOVA, M., WARNOCK, R., DUNBAR, M., MURTHY, P. & DWORKIN, S. L. 2017. Can self-help group programs improve women's empowerment? A systematic review. *Journal of Development Effectiveness,* 9**,** 15-40.

BRUNTON, G., CAIRD, J., KNEALE, D., THOMAS, J. & RICHARDSON, M. 2015a. Community engagement for health via coalitions, collaborations and partnerships: a systematic review and meta-analysis. London: EPPI-Centre, UCL Institute of Education.

BRUNTON, G., OLIVER, S. & THOMAS, J. 2015b. Applying framework synthesis to understand complexity in systematic reviews: A systematic review of methods. *23rd Cochrane Colloquium.* Vienna, Austria.

CAMPBELL COLLABORATION 2017. Protocol Template.

CARR-HILL, R., ROLLESTON, C. & SCHENDEL, R. 2016. The effects of school-based decision-making on educational outcomes in low-and middle-income contexts. *Campbell Systematic Reviews,* 12.

CARTWRIGHT, N. 2007a. Are RCTs the gold standard? *BioSocieties,* 2**,** 11-20.

CARTWRIGHT, N. 2007b. *Hunting causes and using them: Approaches in philosophy and economics*, Cambridge University Press.

CARTWRIGHT, N. 2010. What are randomised controlled trials good for? *Philosophical studies,* 147**,** 59.

CARTWRIGHT, N. 2018. What evidence should guidelines take note of? *Journal of evaluation in clinical practice.*

CHANDLER, J., THOMAS, J., SUTCLIFFE, K., KAHWATI, L. & KNEALE, D. 2017. Applying current philosophical insights on causality using Qualitative Comparative Analysis as an additional synthesis in systematic reviews to address complex interventions. *Gobal Evidence Summit.* Cape Town, South Africa.

CLARK, A. M. 2013. What are the components of complex interventions in healthcare? Theorizing approaches to parts, powers and the whole intervention. *Social Science & Medicine,* 93**,** 185-193.

CLARK, H. & ANDERSON, A. A. 2004. Theories of Change and Logic Models: Telling Them Apart. *American Evaluation Association.* Atlanta, Georgia.

CLARKE, M. & HALSEY, J. 2001. DICE 2: a further investigation of the effects of chance in life, death and subgroup analyses. *International journal of clinical practice,* 55**,** 240-242.

CLARKE, M. & HALSEY, J. 2014. Dicing with chance, life and death in systematic reviews and meta-analyses: DICE 3, a simulation study. *Journal of the Royal Society of Medicine,* 107**,** 116-119.

CLASEN, T. F., BOSTOEN, K., SCHMIDT, W. P., BOISSON, S., FUNG, I. C. H., JENKINS, M. W., SCOTT, B., SUGDEN, S. & CAIRNCROSS, S. 2010. Interventions to improve disposal of human excreta for preventing diarrhoea. *The Cochrane Library*.

CORNISH, F. 2015. Evidence synthesis in international development: a critique of systematic reviews and a pragmatist alternative. *Anthropology & medicine,* 22**,** 263-277.

CRAVER, C. F. 2001. Role functions, mechanisms, and hierarchy. *Philosophy of science,* 68**,** 53-74.

DANGOUR, A. D., WATSON, L., CUMMING, O., BOISSON, S., CHE, Y., VELLEMAN, Y., CAVILL, S., ALLEN, E. & UAUY, R. 2011. Interventions to improve water quality and supply, sanitation and hygiene practices, and their effects on the nutritional status of children. *Cochrane Database Syst Rev,* 3.

DAVIES, P. 2006. What is Needed From Research Synthesis From a Policy Making Perspective? *In:* POPAY, J. (ed.) *Putting Effectiveness Into Context.* London Prime Minister's Strategy Unit, Cabinet Office, United Kingdom.

DAVIES, R. 2018. Representing Theories Of Change: A Technical Challenge With Evaluation Consequences. London: Centre of Excellence for Development Impact and Learning (CEDIL).

DE BUCK, E., HANNES, K., CARGO, M., VAN REMOORTEL, H., VANDE VEEGAETE, A., MOSLER, H.-J., GOVENDER, T., VANDEKERCKHOVE, P. & YOUNG, T. 2018. Engagement of stakeholders in the development of a Theory of Change for handwashing and sanitation behaviour change. *International journal of environmental health research,* 28**,** 8-22.

DE BUCK, E., VAN REMOORTEL, H., HANNES, K., GOVENDER, T., NAIDOO, S., AVAU, B., MUSEKIWA, A., LUTJE, V., CARGO, M. & MOSLER, H.-J. 2017. Approaches to promote handwashing and sanitation behaviour change in low-and middle income countries: a mixed method systematic review. *Campbell Systematic Reviews,* 7**,** 1-447.

DFID. 2017. *DFID 7543 Research Director and Intellectual Leadership - Centre for Excellence for Development Impact and Learning: PO 7543 Contract Section 3 - Terms of Reference* [Online]. London: Department for International Development. [Accessed May 27th 2018].

DIELEMAN, M., GERRETSEN, B. & VAN DER WILT, G. J. 2009. Human resource management interventions to improve health workers' performance in low and middle income countries: a realist review. *Health Research Policy and Systems,* 7**,** 7.

DONALDSON, S. I., STREET, G., SUSSMAN, S. & TOBLER, N. 2001. Using meta-analyses to improve the design of interventions. *In:* SUSSMAN, S. (ed.) *Handbook of program development for health behavior research and practice.* Thousand Oaks, California: Sage.

DUVENDACK, M., HOMBRADOS, J. G., PALMER-JONES, R. & WADDINGTON, H. 2012. Assessing 'what works' in international development: meta-analysis for sophisticated dummies. *Journal of development effectiveness,* 4**,** 456-471.

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.

FUNNELL, S. C. & ROGERS, P. J. 2011. *Purposeful program theory: effective use of theories of change and logic models,* San Francisco, CA, John Wiley & Sons.

GERTLER, P. J., MARTINEZ, S., PREMAND, P., RAWLINGS, L. B. & VERMEERSCH, C. M. 2016. *Impact evaluation in practice,* Washington DC, World Bank Publications.

GLOUBERMAN, S. & ZIMMERMAN, B. 2002. Complicated and complex systems: what would successful reform of Medicare look like? *Commission on the Future of Healthcare in Canada.* Ottawa, Canada.

GOUGH, D., OLIVER, S. & THOMAS, J. 2017. *An introduction to systematic reviews*, Sage.

GOUGH, D. & THOMAS, J. 2017. Commonality and diversity in reviews. *In:* GOUGH, D., OLIVER, S. & THOMAS, J. (eds.) *An Introduction to Systematic Reviews.* London: Sage.

GOUGH, D., THOMAS, J. & OLIVER, S. 2012. Clarifying differences between review designs and methods. *Systematic reviews,* 1**,** 1.

GREENHALGH, J. 2014. Realist synthesis. *In:* EDWARDS, P., O'MAHONEY, J. & VINCENT, S. (eds.) *Studying Organizations Using Critical Realism: A Practical Guide.* Oxford: Oxford University Press.

GREENHALGH, T., ROBERT, G., BATE, P., MACFARLANE, F. & KYRIAKIDOU, O. 2008. *Diffusion of innovations in health service organisations: a systematic literature review*, John Wiley & Sons.

GUYATT, G. H., OXMAN, A. D., VIST, G. E., KUNZ, R., FALCK-YTTER, Y., ALONSO-COELLO, P. & SCHÜNEMANN, H. J. 2008. Rating quality of evidence and strength of recommendations: GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ: British Medical Journal,* 336**,** 924.

HALES, D. 2010. An introduction to triangulation. *UNAIDS monitoring and evaluation*.

HARDEN, A., THOMAS, J., CARGO, M., HARRIS, J., PANTOJA, T., FLEMMING, K., BOOTH, A., GARSIDE, R., HANNES, K. & NOYES, J. 2017. Cochrane Qualitative and Implementation Methods Group Guidance Paper 4: Methods for integrating qualitative and implementation evidence within intervention effectiveness reviews. *Journal of clinical epidemiology*.

HARRIS, K. M., KNEALE, D., LASSERSON, T. J., MCDONALD, V. M., GRIGG, J. & THOMAS, J. 2015. School- based self management interventions for asthma in children and adolescents: a mixed methods systematic review. *The Cochrane Library*.

HARRIS, K. M., KNEALE, D., LASSERSON, T. J., MCDONALD, V. M., GRIGG, J. & THOMAS, J. in press. School-based self management interventions for asthma in children and adolescents: a mixed methods systematic review. *The Cochrane Library*.

HIGGINS, J. P., ALTMAN, D. G., GØTZSCHE, P. C., JÜNI, P., MOHER, D., OXMAN, A. D., SAVOVIĆ, J., SCHULZ, K. F., WEEKS, L. & STERNE, J. A. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *Bmj,* 343**,** d5928.

HIGGINS, J. P. T., LÓPEZ-LÓPEZ, J. A., BECKER, B. J., DAVIES, S. R., DAWSON, S., GRIMSHAW, J. M., MCGUINESS, L. A., MOORE, T. A., REHFUESS, E., THOMAS, J. & CALDWELL, D. M. in press. Synthesising quantitative evidence in systematic reviews of complex health interventions.

HILL, A. B. 1965. The environment and disease: association or causation? *Journal of the Royal Society of Medicine,* 58**,** 295-300.

HOFFMANN, T. C., OXMAN, A. D., IOANNIDIS, J. P., MOHER, D., LASSERSON, T. J., TOVEY, D. I., STEIN, K., SUTCLIFFE, K., RAVAUD, P. & ALTMAN, D. G. 2017. Enhancing the usability of systematic reviews by improving the consideration and description of interventions. *BMJ,* 358**,** j2998.

HOSSAIN, S., SPURWAY, K., ZWI, A. B., HUQ, N. L., MAMUN, R., ISLAM, R., NOWRIN, I., ETHER, S., BONNITCHA, J., DAHAL, N. & ADAMS, A. M. 2017. What is the impact of urbanisation on risk of, and vulnerability to, natural disasters? What are the effective approaches for reducing exposure of urban population to disaster risks? London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

HOWICK, J., GLASZIOU, P. & ARONSON, J. K. 2009. The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation'contribute? *Journal of the Royal Society of Medicine,* 102**,** 186-194.

IEMMI, V., SURESH KUMAR, K., BLANCHET, K., GIBSON, L., HARTLEY, S., MURTHY, G. V., PATEL, V., WEBER, J. & KUPER, H. 2016. Community-based rehabilitation for people with physical and mental disabilities in low-and middle-income countries. *Journal of Development Effectiveness,* 8.

IKA, L. A. & DONNELLY, J. 2017. Success conditions for international development capacity building projects. *International Journal of Project Management,* 35**,** 44-63.

ILLARI, P. & RUSSO, F. 2014. *Causality: Philosophical theory meets scientific practice*, OUP Oxford.

KANE, S. S., GERRETSEN, B., SCHERPBIER, R., DAL POZ, M. & DIELEMAN, M. 2010. A realist synthesis of randomised control trials involving use of community health workers for delivering child health interventions in low and middle income countries. *BMC health services research,* 10**,** 286.

KELLY, M. P. & RUSSO, F. 2017. Causal narratives in public health: the difference between mechanisms of aetiology and mechanisms of prevention in non- communicable diseases. *Sociology of health & illness*.

KNEALE, D., O'MARA-EVES, A. & THOMAS, J. under review. It's a mean world for generalisability: Knowledge claims of generalisability from meta-analyses of public health interventions.

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*.

KNEALE, D., THOMAS, J., O'MARA-EVES, A. & WIGGINS, R. D. 2018. How can additional secondary data analysis of observational data enhance the generalisability of meta-analytic evidence for local public health decision-making? . *Research synthesis methods*.

KRAUER, F., RIESEN, M., REVEIZ, L., OLADAPO, O. T., MARTÍNEZ-VEGA, R., PORGO, T. V., HAEFLIGER, A., BROUTET, N. J., LOW, N. & GROUP, W. Z. C. W. 2017. Zika virus infection as a cause of congenital brain abnormalities and Guillain–Barré syndrome: systematic review. *PLoS medicine,* 14**,** e1002203.

KRIEGER, N. & DAVEY SMITH, G. 2016. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *International journal of epidemiology,* 45**,** 1787-1808.

KRISTJANSSON, E., FRANCIS, D., LIBERATO, S., GREENHALGH, T., WELCH, V., JANDU, M., BATAL, M., RADER, T., NOONAN, E., JANZEN, L., SHEA, B., WELLS, G. A. & PETTICREW, M. 2016. Supplementary feeding for improving the health of disadvantaged infants and children. London: International Initiative for Impact Evaluation (3ie).

LANGER, L., TRIPNEY, J., ERASMUS, Y., TANNOUS, N., CHISORO, C., OPONDO, M., ZIGANA, L., OBUKU, E., VAN ROOYEN, C. & STEWART, R. 2018. Women in wage labour: A systematic review of the effectiveness and design features of interventions supporting women's participation in wage labour in higher growth and/or male-dominated sectors in LMICs. London: Evidence for Policy and Practice Information and Coordinating Centre, Social Science Research Unit, UCL Institute of Education; University College London.

LAWRY, S., SAMII, C., HALL, R., LEOPOLD, A., HORNBY, D. & MTERO, F. 2017. The impact of land property rights interventions on investment and agricultural productivity in developing countries: a systematic review. *Journal of Development Effectiveness,* 9**,** 61-81.

LEWIN, S., BOOTH, A., GLENTON, C., MUNTHE-KAAS, H., RASHIDIAN, A., WAINWRIGHT, M., BOHREN, M. A., TUNÇALP, Ö., COLVIN, C. J. & GARSIDE, R. 2018. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. BioMed Central.

LEWIN, S., HENDRY, M., CHANDLER, J., OXMAN, A. D., MICHIE, S., SHEPPERD, S., REEVES, B. C., TUGWELL, P., HANNES, K. & REHFUESS, E. A. 2017. Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT_SR). *BMC medical research methodology,* 17**,** 76.

LI, T., PUHAN, M. A., VEDULA, S. S., SINGH, S. & DICKERSIN, K. 2011. Network meta-analysis-highly attractive but more methodological research is needed. *BMC medicine,* 9**,** 79.

LUND, C., DE SILVA, M., PLAGERSON, S., COOPER, S., CHISHOLM, D., DAS, J., KNAPP, M. & PATEL, V. 2011. Poverty and mental disorders: breaking the cycle in low-income and middle-income countries. *The Lancet,* 378**,** 1502-1514.

MACKIE, J. L. 1965. Causes and conditions. *American philosophical quarterly,* 2**,** 245-264.

MADEN, M., CUNLIFFE, A., MCMAHON, N., BOOTH, A., CAREY, G. M., PAISLEY, S., DICKSON, R. & GABBAY, M. 2017. Use of programme theory to understand the differential effects of interventions across socio-economic groups in systematic reviews—a systematic methodology review. *Systematic reviews,* 6**,** 266.

MALLETT, R., HAGEN-ZANKER, J., SLATER, R. & DUVENDACK, M. 2012. The benefits and challenges of using systematic reviews in international development research. *Journal of development effectiveness,* 4**,** 445-455.

MAXWELL, J. A. 2004a. Causal explanation, qualitative research, and scientific inquiry in education. *Educational researcher,* 33**,** 3-11.

MAXWELL, J. A. 2004b. Using qualitative methods for causal explanation. *Field methods,* 16**,** 243-264.

MEKASHA, T. J. & TARP, F. 2013. Aid and growth: What meta-analysis reveals. *The journal of development studies,* 49**,** 564-583.

NKOHKWO, Q. N.-A. & ISLAM, M. S. 2013. Challenges to the Successful Implementation of e-Government Initiatives in Sub-Saharan Africa: A Literature Review. *Electronic Journal of e-Government,* 11.

NORRIS, S. L. & BERO, L. 2016. GRADE methods for guideline development: time to evolve? *Annals of internal medicine,* 165**,** 810-811.

NOYES, J., GOUGH, D., LEWIN, S., MAYHEW, A., MICHIE, S., PANTOJA, T., PETTICREW, M., POTTIE, K., REHFUESS, E. & SHEMILT, I. 2013. A research and development agenda for systematic reviews that ask complex questions about complex interventions. *Journal of clinical epidemiology,* 66**,** 1262-1270.

O'MARA-EVES, A., BRUNTON, G., MCDAID, G., OLIVER, S., KAVANAGH, J., JAMAL, F., MATOSEVIC, T., HARDEN, A. & THOMAS, J. 2013. Community engagement to reduce inequalities in health: a systematic review, meta-analysis and economic analysis. *Public Health Research,* 1.

OLIVER, S., BANGPAN, M., STANSFIELD, C. & STEWART, R. 2015. Capacity for conducting systematic reviews in low-and middle-income countries: a rapid appraisal. *Health research policy and systems,* 13**,** 23.

OLIVER, S., GARNER, P., HEYWOOD, P., JULL, J., DICKSON, K., BANGPAN, M., ANG, L., FOURMAN, M. & GARSIDE, R. 2017. Transdisciplinary working to shape systematic reviews and interpret the findings: commentary. *Environmental Evidence,* 6**,** 28.

OLIVER, S., ROCHE, C., STEWART, R., BANGPAN, M., DICKSON, K., PELLS, K., CARTWRIGHT, N., GOUGH, D. & HARGREAVES, J. 2018. Stakeholder Engagement for Development Impact Evaluation and Evidence Synthesis *CEDIL Inception Paper.* London: Centre of Excellence for Development Impact and Learning (CEDIL), London International Development Centre

OYA, C., SCHAEFER, F., SKALIDOU, D., MCCOSKER, C. & LANGER, L. 2017. Effects of certification schemes for agricultural production on socio-economic outcomes in low-and middle-income countries. *Campbell Systematic Reviews,* 13.

PAWSON, R. 2002. Evidence-based policy: The promise of 'realist synthesis'. *Evaluation,* 8**,** 340-358.

PAWSON, R. 2006. *Evidence-based policy: a realist perspective,* Thousand Oaks, CA, Sage.

PAWSON, R., GREENHALGH, T., HARVEY, G. & WALSHE, K. 2005. Realist review-a new method of systematic review designed for complex policy interventions. *Journal of health services research & policy,* 10**,** 21-34.

PAWSON, R. & TILLEY, N. 1997. *Realistic evaluation,* London, Sage.

PETTICREW, M. & ROBERTS, H. 2003. Evidence, hierarchies, and typologies: horses for courses. *Journal of Epidemiology & Community Health,* 57**,** 527-529.

PETTICREW, M., SHEMILT, I., LORENC, T., MARTEAU, T., MELENDEZ-TORRES, G., O'MARA-EVES, A., STAUTZ, K. & THOMAS, J. 2016. Alcohol advertising and public health: systems perspectives versus narrow perspectives. *Journal of epidemiology and community health,* 71**,** 308-312.

PETTICREW, M., TUGWELL, P., KRISTJANSSON, E., OLIVER, S., UEFFING, E. & WELCH, V. 2011. Damned if you do, damned if you don't: subgroup analysis and equity. *Journal of epidemiology and community health,* 66**,** 95-98.

PFADENHAUER, L., ROHWER, A., BURNS, J., BOOTH, A., LYSDAHL, K. B., HOFMANN, B., GERHARDUS, A., MOZYGEMBA, K., TUMMERS, M. & WAHLSTER, P. 2016. Guidance for the Assessment of Context and Implementation in Health Technology Assessments (HTA) and Systematic Reviews of Complex Interventions: The Context and Implementation of Complex Interventions (CICI) Framework.

PLUYE, P. & HONG, Q. N. 2014. Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. *Annual review of public health,* 35.

RAGIN, C. C. 2008. *Redesigning social inquiry: Fuzzy sets and beyond*, Wiley Online Library.

RANIS, G., STEWART, F. & RAMIREZ, A. 2000. Economic growth and human development. *World development,* 28**,** 197-219.

REES, R. & OLIVER, S. 2012. Stakeholder perspectives and participation in reviews. *In:* GOUGH, D., OLIVER, S. & THOMAS, J. (eds.) *An Introduction to Systematic Reviews.* London: Sage Publications.

REHFUESS, E. A., BOOTH, A., BRERETON, L., BURNS, J., GERHARDUS, A., MOZYGEMBA, K., OORTWIJN, W., PFADENHAUER, L. M., TUMMERS, M. & WILT, G. J. 2017. Towards a taxonomy of logic models in systematic reviews and health technology assessments: a priori, staged and iterative approaches. *Research synthesis methods,* 9**,** 13-24.

REISS, J. 2009. Causation in the social sciences: Evidence, inference, and purpose. *Philosophy of the Social Sciences,* 39**,** 20-40.

REISS, J. 2012. Causation in the sciences: An inferentialist account. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences,* 43**,** 769-777.

RILEY, R. D., LAMBERT, P. C. & ABO-ZAID, G. 2010. Meta-analysis of individual participant data: rationale, conduct, and reporting. *Bmj,* 340.

RITCHIE, J. & SPENCER, L. 2002. Qualitative data analysis for applied policy research. *The qualitative researcher's companion,* 573**,** 305-329.

ROGERS, P. 2014. Overview: Strategies for Causal Attribution: Impact Evaluation No. 6. *Methodological Briefs.* Florence, Italy: UNICEF.

ROGERS, P. J. 2000. Causal models in program theory evaluation. *New directions for evaluation,* 2000**,** 47-55.

ROGERS, P. J. 2008. Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation,* 14**,** 29-48.

ROHWER, A., PFADENHAUER, L., BURNS, J., BRERETON, L., GERHARDUS, A., BOOTH, A., OORTWIJN, W. & REHFUESS, E. 2017. Series: Clinical Epidemiology in South Africa. Paper 3: Logic models help make sense of complexity in systematic reviews and health technology assessments. *Journal of clinical epidemiology,* 83**,** 37-47.

RUTTER, H., SAVONA, N., GLONTI, K., BIBBY, J., CUMMINS, S., FINEGOOD, D. T., GREAVES, F., HARPER, L., HAWE, P. & MOORE, L. 2017. The need for a complex systems model of evidence for public health. *The Lancet,* 390**,** 2602-2604.

RYCHETNIK, L., FROMMER, M., HAWE, P. & SHIELL, A. 2002. Criteria for evaluating evidence on public health interventions. *Journal of epidemiology and community health,* 56**,** 119-127.

RYCROFT-MALONE, J., MCCORMACK, B., HUTCHINSON, A. M., DECORBY, K., BUCKNALL, T. K., KENT, B., SCHULTZ, A., SNELGROVE-CLARKE, E., STETLER, C. B. & TITLER, M. 2012. Realist synthesis: illustrating the method for implementation research. *Implementation Science,* 7**,** 33.

SHADISH, W., CAMPBELL, D. T. & COOK, T. D. 2002. *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin Boston.

SHIELL, A., HAWE, P. & GOLD, L. 2008. Complex interventions or complex systems? Implications for health economic evaluation. *BMJ: British Medical Journal,* 336**,** 1281.

SKRONDAL, A. & RABE- HESKETH, S. 2005. *Structural equation modeling: categorical variables*, Wiley Online Library.

SNILSTVEIT, B. 2012. Systematic reviews: from 'bare bones' reviews to policy relevance. *Journal of development effectiveness,* 4**,** 388-408.

SNILSTVEIT, B., OLIVER, S. & VOJTKOVA, M. 2012. Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of development effectiveness,* 4**,** 409-429.

SNILSTVEIT, B., STEVENSON, J., MENON, R., PHILLIPS, D., GALLAGHER, E., GELEEN, M., JOBSE, H., SCHMIDT, T. & JIMENEZ, E. 2016. The impact of education programmes on learning and school participation in low-and middle-income countries. London: 3ie.

SNOW, M. E. & SNOW, N. 2017. Interactive logic models: Using design and technology to explore the effects of dynamic situations on program logic. *Evaluation Journal of Australasia,* 17**,** 20-28.

STEWART, R., LANGER, L., DA SILVA, N. R., MUCHIRI, E., ZARANYIKA, H., ERASMUS, Y., RANDALL, N., RAFFERTY, S., KORTH, M. & MADINGA, N. 2015. The Effects of Training, Innovation and New Technology on African Smallholder Farmers' Wealth and Food Security: A Systematic Review. *Campbell Systematic Reviews,* 11.

SUN, X., BRIEL, M., WALTER, S. D. & GUYATT, G. H. 2010. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *Bmj,* 340**,** c117.

SUTCLIFFE, K., RICHARDSON, M., REES, R., MELENDEZ-TORRES, G., STANSFIELD, C. & THOMAS, J. 2016. What are the critical features of successful Tier 2 weight management programmes. *A systematic review to identify the programme characteristics, and combinations of characteristics, that are associated with successful weight loss. London: EPPI-Centre, UCL*.

SUTCLIFFE, K., THOMAS, J., STOKES, G., HINDS, K. & BANGPAN, M. 2015. Intervention Component Analysis (ICA): a pragmatic approach for identifying the critical features of complex interventions. *Systematic reviews,* 4**,** 1.

THOMAS, J., BRUNTON, J. & GRAZIOSI, S. 2010. EPPI-Reviewer 4.0: software for research synthesis. London: Institute of Education.

THOMAS, J. & HARDEN, A. 2008. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC medical research methodology,* 8**,** 45.

THOMAS, J., HARDEN, A., OAKLEY, A., OLIVER, S., SUTCLIFFE, K., REES, R., BRUNTON, G. & KAVANAGH, J. 2004. Integrating qualitative research with trials in systematic reviews. *BMJ: British Medical Journal,* 328**,** 1010.

THOMAS, J., O'MARA-EVES, A. & BRUNTON, G. 2014. Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example. *Systematic reviews,* 3**,** 1-14.

THOMPSON, S. G. & HIGGINS, J. P. 2005. Can meta-analysis help target interventions at individuals most likely to benefit? *The Lancet,* 365**,** 341-346.

TON, G., DESIERE, S., VELLEMA, W., WYTSE, S. & D'HAESE, M. 2017. The effectiveness of contract farming for raising income of smallholder farmers in low-and middle-income countries: a systematic review. *Campbell Systematic Reviews,* 13.

TRIPNEY, J., HOMBRADOS, J. G., NEWMAN, M., HOVISH, K., BROWN, C., STEINKA-FRY, K. T. & WILKEY, E. 2013. Post-Basic Technical and Vocational Education and Training (TVET) Interventions to Improve Employability and Employment of TVET Graduates in Low-and Middle-Income Countries: A Systematic Review. *Campbell Systematic Reviews,* 9.

TRIPNEY, J., ROULSTONE, A., HOGREBE, N., VIGURS, C., SCHMIDT, E. & STEWART, R. 2015. Interventions to improve the labour market situation of adults with physical and/or sensory disabilities in low-and middle-income countries: a systematic review. *Campbell Systematic Reviews,* 11.

UNICEF 2007. An Evaluation of the PlayPump water system as an appropriate Technology for Water, Sanitation and Hygiene Programmes. New York, USA: United Nations Children's Fund.

VAESSEN, J., RIVAS, A., DUVENDACK, M., PALMER-JONES, R., LEEUW, F. L., VAN GILS, G., LUKACH, R., HOLVOET, N., BASTIAENSEN, J. & HOMBRADOS, J. G. 2014. The effect of microcredit on women's control over household spending in developing countries: a systematic review. *Campbell Systematic Reviews,* 10.

VAN DER KNAAP, L. M., LEEUW, F. L., BOGAERTS, S. & NIJSSEN, L. T. 2008. Combining Campbell standards and the realist evaluation approach. *American journal of evaluation,* 29**,** 48-57.

VAN EIJK, A. M., HILL, J., LARSEN, D. A., WEBSTER, J., STEKETEE, R. W., EISELE, T. P. & TER KUILE, F. O. 2013. Coverage of intermittent preventive treatment and insecticide-treated nets for the control of malaria during pregnancy in sub-Saharan Africa: a synthesis and meta-analysis of national survey data, 2009–11. *The Lancet infectious diseases,* 13**,** 1029-1042.

VANDENBROUCKE, J. P., BROADBENT, A. & PEARCE, N. 2016. Causality and causal inference in epidemiology: the need for a pluralistic approach. *International journal of epidemiology,* 45**,** 1776-1786.

VIGNERI, M., MASSET, E., CLARK, M., EXLEY, J., TUGWELL, P., WELCH, V. & WHITE, H. 2018. Epidemiology and Econometrics: Two Sides of the Same Coin or Different Currencies? *CEDIL Inception Paper.* London: Centre for Excellence in Development, Impact and Learning.

VOLLMER, S., KHAN, S., LE THI NGOC TU, A. & PASHA, S. S. 2017. Protocol: The effect of interventions for women's empowerment on children's health and education: A systematic review of evidence from low-and middle-income countries. Oslo, Norway: Campbell Collaboration.

WADDINGTON, H., MASSET, E. & JIMENEZ, E. 2018. What have we learned after ten years of systematic reviews in international development? *Journal of development effectiveness,* 10**,** 1-16.

WADDINGTON, H. & SNILSTVEIT, B. 2009. Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea. *Journal of development effectiveness,* 1**,** 295-335.

WADDINGTON, H., SNILSTVEIT, B., HOMBRADOS, J. G., VOJTKOVA, M., ANDERSON, J., PHILLIPS, D., DAVIES, P. & WHITE, H. 2014a. Farmer field schools for improving farming practices and farmer outcomes in low-and middle-income countries: a systematic review. *Campbell systematic reviews,* 10.

WADDINGTON, H., WHITE, H. & ANDERSON, J. 2014b. Farmer field schools: From agricultural extension to adult education. *Systematic review summary.* London: International Initiative for Impact Evaluation (3ie).

WADDINGTON, H., WHITE, H., SNILSTVEIT, B., HOMBRADOS, J. G., VOJTKOVA, M., DAVIES, P., BHAVSAR, A., EYERS, J., KOEHLMOOS, T. P. & PETTICREW, M. 2012. How to do a good systematic review of effects in international development: a tool kit. *Journal of development effectiveness,* 4**,** 359-387.

WEED, D. L. 2000. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *International Journal of Epidemiology,* 29**,** 387-390.

WELCH, V. A., GHOGOMU, E., HOSSAIN, A., AWASTHI, S., BHUTTA, Z. A., CUMBERBATCH, C., FLETCHER, R., MCGOWAN, J., KRISHNARATNE, S. & KRISTJANSSON, E. 2017. Mass deworming to improve developmental health and wellbeing of children in low-income and middle-income countries: a systematic review and network meta-analysis. *The Lancet Global Health,* 5**,** e40-e50.

WESTHORP, G., WALKER, B., ROGERS, P., OVERBEEKE, N., BALL, D. & BRICE, G. 2014. Enhancing community accountability, empowerment and education outcomes in low and middle-

income countries: A realist review. *University of London, Institute of Education, EPPI-Centre, London*.

WHITE, H. 2009. Theory-based impact evaluation: principles and practice. *Journal of development effectiveness,* 1**,** 271-284.

WHITE, H. 2014. Current challenges in impact evaluation. *The European Journal of Development Research,* 26**,** 18-30.

WHITE, H. 2018. Theory-based systematic reviews. *Journal of Development Effectiveness,* 10**,** 17-38.

WHITESIDE, M. F. & BECKER, B. J. 2000. Parental factors and the young child's postdivorce adjustment: a meta-analysis with implications for parenting arrangements. American Psychological Association.

WILLEY, B., SMITH PAINTAIN, L., MANGHAM, L., CAR, J. & ARMSTRONG SCHELLENBERG, J. 2013. Effectiveness of interventions to strengthen national health service delivery on coverage, access, quality and equity in the use of health services in low and lower middle income countries. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London

WIRTZ, P. W. 2007. Advances in causal chain development and testing in alcohol research: Mediation, suppression, moderation, mediated moderation, and moderated mediation. *Alcoholism: Clinical and Experimental Research,* 31.

WOOLDRIDGE, J. M. 2015. *Introductory econometrics: A modern approach*, Nelson Education.

Contact Us

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