

Conceptualising causal pathways in systematic reviews of international development interventions through adopting a causal chain analysis approach

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 logic models as a tool for outlining graphically a hypothesis of how an intervention leads to a change in an outcome through depicting a causal chain of events. However, it is the nature of these connecting relationships and their basis in causality which is of interest here, and we focus on complex causal relationships and the way in which contextual factors reflecting the intervention setting or population may moderate these. Evidence synthesis techniques are considered, and their usefulness in analysing 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.

Keywords: evidence synthesis, causal chain analysis, logic models, systematic reviews, causality

Introduction

Causal chain analysis (CCA) 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 a plurality of synthesis methods. 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 graphical form (Kneale, Thomas, and Harris 2015; Anderson et al. 2011). 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 which types of synthesis methods are most appropriate. This includes thinking through what types of causal claims we are making, and what types of causal account we are drawing upon in doing so. 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. The overall aim of this paper is to guide reviewers in undertaking CCA particularly around issues of causality. We explore how logic models are developed and the types of assumptions and relationships that are depicted, as well as briefly the type of synthesis method(s) that can be used to explore different types of relationship; readers may also wish to consult a more detailed report published elsewhere simultaneously (Kneale, Thomas, Bangpan, et al. 2018).

Causal Thinking and Systematic Reviews

Well-conducted systematic reviews begin with a clearly defined research question and an articulation of the conceptual framework (Gough, Oliver, and Thomas 2017; 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. 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:

- (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).
- (ii) **Probabilistic accounts** arise from statistical analyses of quantitative data (Reiss 2009). Many probabilistic accounts of causal relationships are based on classical linear regression models (Reiss 2009), or extensions to these, which aim to model the effect of a *ceteris paribus* change (all other factors being equal) in one variable (intervention exposure) on another (outcome). Probabilistic accounts of causality have been described as indeterministic or stochastic, in that they indicate broad-brushed trends, for example at a population level, but random variation and un/observed 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, 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 sub-population (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, 2010; Illari and Russo 2014). However, this extrapolation can be problematic for a number of reasons (see Kneale, Thomas, O'Mara-Eves, et al. 2018).

(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, O'Mara-Eves, and Brunton 2014; Cartwright 2007b; Reiss 2009), but are interpreted using mechanistic reasoning in systematic reviews.

(iv) **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, Broadbent, and Pearce 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. 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).

- (v) **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 of the intervention) and activities (what these entities do) are organised to effect a change (mechanism) in the outcome(s) (Illari and Russo 2014). 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).

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, 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). 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. Understanding our own epistemological standpoint is important when it comes to evaluating the 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, and those that ‘merely vouch’ for their conclusions, for example QCA (Cartwright 2007b). Statements involving ‘clinchers’ suggest that if the underlying assumptions are met, a positive result implies causality and clinches the conclusion (for example evidence from randomised controlled trials (RCTs)), rather than merely vouches for it (Cartwright 2007b, 2007a). In contrast evidence that ‘vouches’ for a causal relationship ‘speaks’ of a causal relationship, but other evidence is needed to secure the conclusion of causality (Cartwright 2018, , p6). Cartwright highlights the weakness in terms of applicability of the former, and the uncertainty (and potential bias) surrounding the latter form of causal claim.

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, Thomas, and Harris 2015), and the identification of suitable synthesis methods to support exploration of the model.

Making Links Between Interventions and Outcomes

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:

1. The intervention component
2. The function or purpose of the component
3. The output or outcome it is intended to change
4. 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 mechanisms, 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 where the whole intervention leads to an outcome that individual component parts cannot lead to (Rogers 2008). Different types of complex and non-linear mechanisms are described in Box 1.

Box 1 – A taxonomy of complex mechanisms

Virtuous circles/cycles (and vicious circles/cycles): A ‘virtuous circle’ is activated when initial changes in the outcome creates the opportunities for further self-reinforcing changes (Rogers 2008). The converse, vicious cycles, are self-reinforcing negative intervention effects.

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, Hawe, and Gold 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.

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. In contrast moderators and interaction effects refer to factors that can amplify or dampen the relationship between exposure to the intervention and the outcomes.

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 of intervention characteristics, each consisting of different components/characteristics which alone cannot trigger an outcome, lead to the same outcome (i.e. multiple complex pathways).

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).

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. These forms of sufficient causal relationships are usually the target of systematic reviews.

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 starting a house fire, as there are many other routes through which homes catch fire.

Determining whether an intervention should be regarded as simple, complicated or complex can reflect the degree to which context is critical for the likely outcome achieved and hence ability to project generalisations from one setting to the next; 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). (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). 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 therefore it is not clear to what extent development interventions are complicated or complex problems. One distinction might be drawn between micro-level problems, like the effectiveness of early child development programmes to improve nutrition (where a review by Kristjansson et al. (2016) stressed the importance of implementation factors, suggesting a more complicated intervention), on the one hand, and macro-level problems, like how to achieve sustainable economic growth at the country-level (where a review by Hausmann, Rodrik, and Velasco (2008) stressed the importance of context in determining economic growth, suggesting a more complex intervention) on the other. Alternatively, 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.

Evidence of Causality in Systematic Reviews Employing Causal Chain Analyses

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. While systematic reviews of RCTs, particularly those that employ meta-analyses, hold potential for developing ‘clincher’ claims of causality (Cartwright 2007b), without seeking mechanistic accounts, we may be less certain why, or how, outcomes are achieved, and the generalisability of

the evidence to situations beyond those included in the meta-analysis. Overcoming these challenges and producing mechanistic accounts of causality often necessitates working with a more diverse set of studies, and criteria or principles for theorising when a relationship is causal can be particularly useful in this case. Some of these are described below in Box 2, although many of the approaches included also address questions around the quality of the studies (focussed on internal validity), and not all of the items included in a tool in some may be suitable for assessing evidence from development interventions.

Box 2: 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, some elements are 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 a body of evidence (usually based on a review of RCT studies), 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). CERQUAL encourages reviewers to consider: (1) methodological limitations in the included studies, (2) coherence, (3) adequacy of data, and (4) relevance. This tool is being used in Cochrane Qualitative Reviews and, as is the case for GRADE (above), is used to assess the quality of a body of evidence, as opposed to assessing individual studies.

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); (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 is 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, Glasziou, and Aronson 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).

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 (2004, 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, as maximising the key strengths of qualitative research for causal attribution. Further details are also outlined in Kneale, Thomas, Bangpan, et al. (2018).

Conceptualising the way in which interventions operate within systems

“Theories of change” and “logic models” 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. While the use, and particularly effective and extensive use, of these techniques within systematic reviews is still at its infancy (Kneale, Thomas, and Harris 2015), 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 effect a change in outcomes, and they can be used to depict the complex mechanisms outlined earlier. The terms “theories of change” and “logic models” are often used interchangeably by reviewers, largely dependent on disciplinary preference, although a subtle distinction exists (see (Kneale, Thomas, Bangpan, et al. 2018)). Within the methodological literature on systematic reviewing, ‘logic model’ has emerged as the favoured terminology, hence we use the term hereon to describe the depiction of causal chains in reviews.

What Does a Logic Model Look Like?

Several examples of logic models exist in the systematic review literature (see (Kneale, Thomas, and Harris 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. 2014, , 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 (Waddington et al. 2014). The 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. 2014). 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. 2014). The authors returned to the model in the final analysis drawing on the evidence synthesised, and articulated different causal chains relating to empowerment, and also showed that the hypothesised virtuous circle depicted in the original logic model was not supported by the evidence.

FIGURE 1 HERE

The steps taken in developing a logic model afresh have been outlined in detail elsewhere (Kneale, Thomas, and Harris 2015) and 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 (for example Anderson et al. 2011; Rohwer et al. 2017; Kneale, Thomas, Bangpan, et al. 2018). Despite the utility of using a logic model to theorise how an intervention works, and its role in supporting subsequent review decisions, it is important to remember that a logic model is a model and entirely new ways of thinking about the intervention may emerge from the evidence

synthesised during the review. It is expected that the logic model will develop on the basis of new understandings brought by evidence synthesised within the systematic review (Kneale, Thomas, and Harris 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. 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.

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). 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). Logic models provide a useful starting point encouraging systematic reviewers to consider 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. Building up an ‘isolated description’ of an intervention’s causal chain (Illari and Russo 2014) 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) may be ultimately more useful for reviews of complex development interventions.

How can different evidence synthesis methods be used to explore causal relationships in systematic reviews?

Many different forms of synthesis can be used to understand and evidence causal chains, and several other sources provide a detailed account of these (Gough, Oliver, and Thomas 2017; Snilstveit, Oliver, and Vojtkova 2012; Waddington et al. 2012). These include narrative synthesis (Snilstveit, Oliver, and Vojtkova 2012), many different forms of qualitative evidence synthesis (Thomas and Harden 2008), as well as different forms of quantitative analysis and meta-analysis (Higgins et al. in press). Those discussed here are featured as they either show promise for causal chain analysis but are comparatively novel in systematic reviews (for instance QCA (Thomas, O'Mara-Eves, and Brunton 2014)); or, as in the case of meta-analysis, there is further scope for their repurposing to undertake CCA.

Exploring causal chains using meta-analysis

Meta-analysis used in CCA involves the quantitative synthesis of data on the direction, magnitude and precision of the impact of interventions on outcomes. A principle of causal chain analysis (CCA) is that complex interventions cannot be understood as a single undifferentiated “whole”. However, many examples of meta-analysis tend to model interventions as binary exposures, lumping together all processes, and clumping all outcomes as changes that occur simultaneously. 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 techniques are, however, repeatedly cautioned that associations observed through such analyses are observational in nature and offer no basis for assuming causality (Thompson and Higgins 2005) and may be subject to many of the same caveats of observational research. 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 be incorporated into meta-analysis and meta-regression models (Waddington et al. 2012).

Extensions to Traditional 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 compare the effectiveness of hypothesised causal chains. An example in literature is a comparison of different approaches to mass deworming interventions and their impact on developmental health and wellbeing 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 cycles that were expected to operate, and then tested comparative effectiveness of different 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). Nevertheless, despite their potential promise, some of the underlying assumptions of network meta-analysis may be difficult to substantiate and may require additional considerations (see Wilson, Tanner-Smith, and Mavridis 2016).

Other extensions to meta-analysis can also help to mirror some of the complexities in hypothesised causal chains and 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) which allows more flexible, complex statistical analysis of study data. This can include observing the extent of missing data, ensuring consistency in modelling strategy and choice of outcomes and confounders, and constructing models that better account for any correlations between outcomes from the same participants or other forms of clustering in the data (see (Riley, Lambert, and Abo-Zaid 2010)). However, at the time of writing, use of IPD meta-analyses remains scarce in the literature, reflecting the paucity of IPD from interventions that are routinely made available to meta-analysts (e.g. through data repositories or shared more directly by triallists). This may change in future if calls for greater availability of IPD data, and repositories that support increased access to IPD data, are successful (Sankoh et al. 2013). Consequently, current examples of reviews that employ IPD meta-analyses in the field of international development are relatively rare (see Purgato et al. (2018) for an

example), and are often 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, 2009). As Becker outlines, unlike some of the more traditional approaches to meta-analysis, 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 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 (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 model-based meta-analyses have been conducted with other data types (e.g. categorical data). In the absence of either IPD data, or sufficiently rich data to support model-based meta-analysis, systematic reviewers may need to turn to 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).

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 has been used as a technique for amalgamating diverse data from quantitative and qualitative studies and for studying complex interventions (Brunton, Oliver, and Thomas 2015). Framework synthesis involves working through defined analytical stages, and can involve deductive phases, where data are synthesised according to the framework, and inductive phases, where evidence that doesn't fit into the framework are considered (Booth and Carroll 2015). 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 causal clinch (Cartwright 2007b).

An 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). A framework specifically for assessing causal relationships between Zika and adverse child outcomes and nervous disorders was developed. A systematic review was then conducted to assess the validity of the framework synthesising evidence from '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 was not explicitly described as 'framework synthesis', the description provided appeared to

encompass several stages of framework synthesis. Through developing and testing a framework using hypothetico-deductive means, and evaluation by an expert panel, the authors concluded that Zika virus was a cause of congenital abnormalities and a trigger of Guillain-Barré syndrome.

Framework synthesis is an attractive method for causal chain analysis as it accommodates the synthesis of different types of evidence which may reflect different strands of the causal chain, and is closely aligned with mechanistic accounts. However, given that it remains a relatively nascent method, its utility is still being realised, and caveats do apply, particularly around the need to develop standards for practice for the conduct of framework synthesis.

Theory-Based Systematic Reviews

Theory based systematic reviews (Snilstveit 2012; Waddington et al. 2014; Waddington et al. 2012; White 2009, 2018), combine 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. Theory-based reviews use logic models, theories of change and occasionally middle-range or high-level (academic) theory. Some theory-based systematic reviews also have a theory-building component in which an initial theory presented at protocol stage is further refined during the synthesis stage as a result of the evidence collection (see Waddington et al. 2014 for an example).

The approach uses explicit theory to shape the whole systematic review process, and a theory-based systematic review may also explicitly attempt 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 drawing on a mix of quantitative and qualitative evidence

(see (White 2018) for further information and guidance). Theory-based systematic reviews are probably most closely related to framework synthesis in the approach to combining broad evidence along the causal chain. A key distinction appears to be that theory-based systematic reviews usually include a research question that focuses on synthesising results from impact evaluations (see Waddington, Masset, and Jimenez (2018)). As with CCA, theory-based systematic reviews are a significant undertaking, and are conceptually challenging, and their production may necessitate drawing upon different causal accounts, which diverge from the traditional counterfactual accounts usually synthesised in (standard) reviews of impact evaluations (White 2018).

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, O'Mara-Eves, and Brunton 2014). QCA is being used within systematic reviews both to further understand the results of meta-analyses (for example Brunton et al. 2015), to develop theories to test within meta-analyses (for example Harris et al. 2015), and occasionally as a synthesis method in its own right. QCA allows us to test causal conditions using a regularity account of causality, albeit with mechanistic interpretation. 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 of two aspects of set relationships, necessity and sufficiency and, building from these, can be used to investigate other complex relationships including multiple conjunctural causation and INUS relationships (described earlier).

An application of QCA in reviews of international development interventions can be explored in Ton et al. (2017)'s review of the evidence on contract farming (which involves establishing a relationship between a farmer and a firm before production begins). 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 included in 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. Another recent application, focussed on increasing female labour market participation, can be found in a recent review by Langer et al. (2018). While promising for the synthesis of complex interventions, some drawbacks of QCA need to be considered, including potential issues in the replicability of the analysis given it involves iterative stages, and the type of (replicability) causal account that is developed during synthesis, which may not sufficiently 'clinch' causality for some decision-makers.

Using Realist Approaches to explore causal chains

Realist synthesis aims to unpack the complexity of programme theory with the concept of a "generative approach to causation", adapted by realist reviewers, signifying that causal mechanisms, rather than "programmes", are the unit of analysis and key to generate desired changes (Pawson 2002). 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-makers and practitioners in 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). As outlined by Pawson (2002), synthesising evidence 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 uncover how programmes lead to (un)expected changes.

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 on 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 which would provide essential information for establishing potential inferential claims. In a review undertaken by Eddy-Spicer et al. (2016) on school accountability systems an initial rough theory was developed at the scoping exercise stage, involving further consultation with advisory group members. Synthesis was an iterative process involving

five rounds of data synthesis, while the final round consisted of a comparison across all school accountability elements (assessment, monitoring, and inspection) applying constant comparative methods to consider mechanisms and make inferential claims. An example finding included: “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” (Eddy-Spicer et al. 2016, , p5-6). Several further examples from international development reviews are included in Kneale, Thomas, Bangpan, et al. (2018), which also cautions that while diverse data can be synthesised in the identification of CMO configurations, the presentation of results is usually narrative or tabular, and quantifiable estimates of a mechanism and its impact is not always clear.

Conclusion

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. 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, Oliver, and Thomas 2017; Thomas and Harden 2008). However, no one synthesis method alone is likely to provide a complete mechanistic account of interventions and analyses of causal

chains require a plurality of forms of evidence; this is in much the same way that philosophers advocate that “evidential pluralism” strengthen causal hypotheses (Krieger and Davey Smith 2016; Reiss 2009, 2012). Alongside selecting an appropriate synthesis method, reviewers need to understand the types of causal account they are developing, the types of causal relationships that are depicted within their logic models, and how they will evaluate causal relationships, in order to interpret and present their findings to decision-makers around the strength of causal claims being made.

Sophisticated statistical modelling to derive an intervention's true impact and estimate its cost-effectiveness 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 provides the “clincher” (Cartwright 2007b), but in such narrow terms that its application elsewhere is challenging. For complex international development interventions, “clinchers” become of limited value for decision-making without understanding the underlying processes. Rogers’ arguments that ‘the art of dealing with the complicated and complex real world lies in knowing when to simplify and when, and how, to complicate’ is useful here (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 understandings of how interventions effect change.

References

- Anderson, Laurie M., Mark Petticrew, Eva Rehfuss, Rebecca Armstrong, Erin Ueffing, Phillip Baker, Daniel Francis, and Peter Tugwell. 2011. "Using logic models to capture complexity in systematic reviews." Review of. *Research synthesis methods* 2 (1):33-42.

- Becker, Betsy Jane. 2001. "Examining theoretical models through research synthesis: The benefits of model-driven meta-analysis." Review of. *Evaluation & the Health Professions* 24 (2):190-217.
- . 2009. "Model-based meta-analysis." In *The handbook of research synthesis and meta-analysis*, edited by Harris Cooper, Larry V. Hedges and Jeffrey C Valentine, 377-95. New York: Russell Sage Foundation.
- Booth, Andrew, and Christopher Carroll. 2015. "How to build up the actionable knowledge base: the role of 'best fit' framework synthesis for studies of improvement in healthcare." Review of. *BMJ Qual Saf* 24 (11):700-8.
- Brunton, Ginny, Jenny Caird, Dylan Kneale, James Thomas, and Michelle Richardson. 2015. "Community engagement for health via coalitions, collaborations and partnerships: a systematic review and meta-analysis." In. London: EPPI-Centre, UCL Institute of Education.
- Brunton, Ginny, Sandy Oliver, and James Thomas. 2015. "Applying framework synthesis to understand complexity in systematic reviews: A systematic review of methods." In *23rd Cochrane Colloquium*. Vienna, Austria.
- Campbell Collaboration. 2017. "Protocol Template." In.
- Cartwright, Nancy. 2007a. "Are RCTs the gold standard?" Review of. *BioSocieties* 2 (1):11-20.
- . 2007b. *Hunting causes and using them: Approaches in philosophy and economics*: Cambridge University Press.
- . 2010. "What are randomised controlled trials good for?" Review of. *Philosophical studies* 147 (1):59.
- . 2018. "What evidence should guidelines take note of?" Review of. *Journal of evaluation in clinical practice*.
- Chandler, Jackie, James Thomas, Katy Sutcliffe, Leila Kahwati, and Dylan Kneale. 2017. "Applying current philosophical insights on causality using Qualitative Comparative Analysis as an additional synthesis in systematic reviews to address complex interventions." In *Global Evidence Summit*. Cape Town, South Africa.
- Clark, Alexander M. 2013. "What are the components of complex interventions in healthcare? Theorizing approaches to parts, powers and the whole intervention." Review of. *Social science & medicine* 93:185-93.
- Eddy-Spicer, David, Melanie Ehren, Mukdarut Bangpan, M Khatwa, and F Perrone. 2016. "Under what conditions do inspection, monitoring and assessment improve system efficiency, service delivery and learning outcomes for the poorest and most marginalised? A realist synthesis of school accountability in low-and middle-income countries." In.: London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.
- Funnell, Sue C., and Patricia J. Rogers. 2011. *Purposeful program theory: effective use of theories of change and logic models*. Vol. 31. San Francisco, CA: John Wiley & Sons.
- Glouberman, Sholom, and Brenda Zimmerman. 2002. "Complicated and complex systems: what would successful reform of Medicare look like?" In *Commission on the Future of Healthcare in Canada*. Ottawa, Canada.
- Gough, David, Sandy Oliver, and James Thomas. 2017. *An introduction to systematic reviews*: Sage.
- Guyatt, Gordon H, Andrew D Oxman, Gunn E Vist, Regina Kunz, Yngve Falck-Ytter, Pablo Alonso-Coello, and Holger J Schünemann. 2008. "Rating quality of evidence and strength of recommendations: GRADE: an emerging consensus on

- rating quality of evidence and strength of recommendations." Review of. *BMJ: British Medical Journal* 336 (7650):924.
- Harris, Katherine M, Dylan Kneale, Toby J Lasserson, Vanessa M McDonald, Jonathan Grigg, and James Thomas. 2015. "School-based self management interventions for asthma in children and adolescents: a mixed methods systematic review." Review of. *The Cochrane Library*. doi: 10.1002/14651858.CD011651.
- Hausmann, Ricardo, Dani Rodrik, and Andrés Velasco. 2008. "Growth diagnostics." In *The Washington consensus reconsidered: Towards a new global governance*, edited by Narcís Serra and Joseph E Stiglitz. Oxford, UK: Oxford University Press
- Higgins, Julian P. T., José A López-López, Betsy J Becker, Sarah R Davies, Sarah Dawson, Jeremy M Grimshaw, Luke A McGuinness, et al. in press. "Synthesising quantitative evidence in systematic reviews of complex health interventions." Review of.
- Hill, Austin Bradford. 1965. "The environment and disease: association or causation?" Review of. *Journal of the Royal Society of Medicine* 58:295-300.
- Howick, Jeremy, Paul Glasziou, and Jeffrey K Aronson. 2009. "The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute?" Review of. *Journal of the Royal Society of Medicine* 102 (5):186-94.
- Illari, Phyllis, and Federica Russo. 2014. *Causality: Philosophical theory meets scientific practice*. OUP Oxford.
- Kane, Sumit S, Barend Gerretsen, Robert Scherpbier, Mario Dal Poz, and Marjolein Dieleman. 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." Review of. *BMC health services research* 10 (1):286.
- Kneale, Dylan, James Thomas, Mukdarut Bangpan, Ian Shemilt, Hugh Waddington, and David Gough. 2018. "Causal chain analysis in systematic reviews of international development interventions." In *CEDIL Inaugural Papers*. London: Centre of Excellence for Development Impact and Learning.
- Kneale, Dylan, James Thomas, and Katherine Harris. 2015. "Developing and Optimising the Use of Logic Models in Systematic Reviews: Exploring Practice and Good Practice in the Use of Programme Theory in Reviews." Review of. *PloS one*. doi: DOI: 10.1371/journal.pone.0142187.
- Kneale, Dylan, James Thomas, Alison O'Mara-Eves, and Richard D. Wiggins. 2018. "How can additional secondary data analysis of observational data enhance the generalisability of meta-analytic evidence for local public health decision-making? ." Review of. *Research synthesis methods*. doi: <https://doi.org/10.1002/jrsm.1320>.
- Krauer, Fabienne, Maurane Riesen, Ludovic Reveiz, Olufemi T Oladapo, Ruth Martínez-Vega, Teegwende V Porgo, Anina Haefliger, Nathalie J Broutet, Nicola Low, and WHO Zika Causality Working Group. 2017. "Zika virus infection as a cause of congenital brain abnormalities and Guillain–Barré syndrome: systematic review." Review of. *PLoS medicine* 14 (1):e1002203.
- Krieger, Nancy, and George Davey Smith. 2016. "The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology." Review of. *International journal of epidemiology* 45 (6):1787-808.
- Kristjansson, Elizabeth, Damian Francis, S Liberato, Trisha Greenhalgh, Vivian Welch, MB Jandu, M Batal, et al. 2016. "Supplementary feeding for improving the

- health of disadvantaged infants and children." In. London: International Initiative for Impact Evaluation (3ie).
- Langer, Laurenz, Janice Tripney, Y Erasmus, N Tannous, C Chisoro, M Opondo, L Zigana, E Obuku, C van Rooyen, and Ruth Stewart. 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." In. London: Evidence for Policy and Practice Information and Coordinating Centre, Social Science Research Unit, UCL Institute of Education; University College London.
- Lewin, Simon, Andrew Booth, Claire Glenton, Heather Munthe-Kaas, Arash Rashidian, Megan Wainwright, Meghan A Bohren, Özge Tunçalp, Christopher J Colvin, and Ruth Garside. 2018. "Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series." In.: BioMed Central.
- Lewin, Simon, Maggie Hendry, Jackie Chandler, Andrew D Oxman, Susan Michie, Sasha Shepperd, Barnaby C Reeves, Peter Tugwell, Karin Hannes, and Eva A Rehfuess. 2017. "Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT_SR)." Review of. *BMC Medical Research Methodology* 17 (1):76.
- Mackie, John L. 1965. "Causes and conditions." Review of. *American philosophical quarterly* 2 (4):245-64.
- Mallett, Richard, Jessica Hagen-Zanker, Rachel Slater, and Maren Duvendack. 2012. "The benefits and challenges of using systematic reviews in international development research." Review of. *Journal of development effectiveness* 4 (3):445-55.
- Maxwell, Joseph A. 2004. "Causal explanation, qualitative research, and scientific inquiry in education." Review of. *Educational Researcher* 33 (2):3-11.
- Norris, Susan L, and Lisa Bero. 2016. "GRADE methods for guideline development: time to evolve?" Review of. *Annals of internal medicine* 165 (11):810-1.
- Pawson, Ray. 2002. "Evidence-based policy: The promise of 'realist synthesis'." Review of. *Evaluation* 8 (3):340-58.
- Pawson, Ray, and Nick Tilley. 1997. *Realistic evaluation*. London: Sage.
- Purgato, Marianna, Alden L Gross, Theresa Betancourt, Paul Bolton, Chiara Bonetto, Chiara Gastaldon, James Gordon, Paul O'Callaghan, Davide Papola, and Kirsi Peltonen. 2018. "Focused psychosocial interventions for children in low-resource humanitarian settings: a systematic review and individual participant data meta-analysis." Review of. *The Lancet Global Health* 6 (4):e390-e400.
- Ragin, Charles C. 2008. *Redesigning social inquiry: Fuzzy sets and beyond*. Vol. 240: Wiley Online Library.
- Reiss, Julian. 2009. "Causation in the social sciences: Evidence, inference, and purpose." Review of. *Philosophy of the Social Sciences* 39 (1):20-40.
- . 2012. "Causation in the sciences: An inferentialist account." Review of. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (4):769-77.
- Riley, Richard D., Paul C. Lambert, and Ghada Abo-Zaid. 2010. "Meta-analysis of individual participant data: rationale, conduct, and reporting." Review of. *Bmj* 340.
- Ritchie, Jane, and Liz Spencer. 2002. "Qualitative data analysis for applied policy research." Review of. *The qualitative researcher's companion* 573 (2002):305-29.

- Rogers, Patricia. 2014. "Overview: Strategies for Causal Attribution: Impact Evaluation No. 6." In *Methodological Briefs*. Florence, Italy: UNICEF.
- Rogers, Patricia J. 2000. "Causal models in program theory evaluation." Review of. *New directions for evaluation* 2000 (87):47-55.
- Rogers, Patricia J. 2008. "Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions." Review of. *Evaluation* 14 (1):29-48.
- Rohwer, Anke, Lisa Pfadenhauer, Jacob Burns, Louise Brereton, Ansgar Gerhardus, Andrew Booth, Wija Oortwijn, and Eva Rehfuss. 2017. "Series: Clinical Epidemiology in South Africa. Paper 3: Logic models help make sense of complexity in systematic reviews and health technology assessments." Review of. *Journal of clinical epidemiology* 83:37-47.
- Rutter, Harry, Natalie Savona, Keteven Glonti, Jo Bibby, Steven Cummins, Diane T Finegood, Felix Greaves, Laura Harper, Penelope Hawe, and Laurence Moore. 2017. "The need for a complex systems model of evidence for public health." Review of. *The Lancet* 390 (10112):2602-4.
- Sankoh, Osman, Abraham J Herbst, Sanjay Juvekar, Stephen Tollman, Peter Byass, and Marcel %J The lancet global health Tanner. 2013. "INDEPTH launches a data repository and INDEPTHStats." Review of. 1 (2):e69.
- Shiell, Alan, Penelope Hawe, and Lisa Gold. 2008. "Complex interventions or complex systems? Implications for health economic evaluation." Review of. *BMJ: British Medical Journal* 336 (7656):1281.
- Snilstveit, Birte. 2012. "Systematic reviews: from 'bare bones' reviews to policy relevance." Review of. *Journal of development effectiveness* 4 (3):388-408.
- Snilstveit, Birte, Sandy Oliver, and Martina Vojtkova. 2012. "Narrative approaches to systematic review and synthesis of evidence for international development policy and practice." Review of. *Journal of development effectiveness* 4 (3):409-29.
- Thomas, James, and Angela Harden. 2008. "Methods for the thematic synthesis of qualitative research in systematic reviews." Review of. *BMC Medical Research Methodology* 8 (1):45.
- Thomas, James, Alison O'Mara-Eves, and Ginny Brunton. 2014. "Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example." Review of. *Systematic reviews* 3 (1):1-14.
- Thompson, Simon G, and Julian PT Higgins. 2005. "Can meta-analysis help target interventions at individuals most likely to benefit?" Review of. *The Lancet* 365 (9456):341-6.
- Ton, Giel, Sam Desiere, Wytse Vellema, Sophia Wytse, and Marijke D'Haese. 2017. "The effectiveness of contract farming for raising income of smallholder farmers in low-and middle-income countries: a systematic review." Review of. *Campbell Systematic Reviews* 13. doi: DOI: 10.4073/csr.2017.13.
- van Eijk, Anna Maria, Jenny Hill, David A Larsen, Jayne Webster, Richard W Steketee, Thomas P Eisele, and Feiko O ter Kuile. 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." Review of. *The Lancet infectious diseases* 13 (12):1029-42.
- Vandenbroucke, Jan P, Alex Broadbent, and Neil Pearce. 2016. "Causality and causal inference in epidemiology: the need for a pluralistic approach." Review of. *International journal of epidemiology* 45 (6):1776-86.

- Waddington, Hugh, Edoardo Masset, and Emmanuel Jimenez. 2018. "What have we learned after ten years of systematic reviews in international development?" Review of. *Journal of development effectiveness* 10 (1):1-16.
- Waddington, Hugh, Birte Snilstveit, Jorge Garcia Hombrados, Martina Vojtkova, Jock Anderson, Daniel Phillips, Philip Davies, and Howard White. 2014. "Farmer field schools for improving farming practices and farmer outcomes in low-and middle-income countries: a systematic review." Review of. *Campbell Systematic Reviews* 10 (6). doi: DOI: 10.4073/csr.2014.6.
- Waddington, Hugh, Howard White, Birte Snilstveit, Jorge Garcia Hombrados, Martina Vojtkova, Philip Davies, Ami Bhavsar, John Eysers, Tracey Perez Koehlmoos, and Mark Petticrew. 2012. "How to do a good systematic review of effects in international development: a tool kit." Review of. *Journal of development effectiveness* 4 (3):359-87.
- Welch, Vivian A, Elizabeth Ghogomu, Alomgir Hossain, Shally Awasthi, Zulfiqar A Bhutta, Chisa Cumberbatch, Robert Fletcher, Jessie McGowan, Shari Krishnaratne, and Elizabeth Kristjansson. 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." Review of. *The Lancet Global Health* 5 (1):e40-e50.
- Westhorp, Gill, Bill Walker, Patricia Rogers, Nathan Overbeeke, Daniel Ball, and Graham Brice. 2014. "Enhancing community accountability, empowerment and education outcomes in low and middle-income countries: A realist review." Review of. *University of London, Institute of Education, EPPI-Centre, London*.
- White, Howard. 2009. "Theory-based impact evaluation: principles and practice." Review of. *Journal of development effectiveness* 1 (3):271-84.
- . 2014. "Current challenges in impact evaluation." Review of. *The European Journal of Development Research* 26 (1):18-30.
- . 2018. "Theory-based systematic reviews." Review of. *Journal of development effectiveness* 10 (1):17-38.
- Whiteside, Mary F, and Betsy Jane Becker. 2000. "Parental factors and the young child's postdivorce adjustment: a meta-analysis with implications for parenting arrangements." In.: American Psychological Association.
- Wilson, David B, Emily Tanner-Smith, and Dimitris Mavridis. 2016. "Campbell Methods Policy Note on Network Meta-Analysis (Version 1.0, updated September 2015)." In. Oslo: The Campbell Collaboration.
- Wirtz, Philip W. 2007. "Advances in causal chain development and testing in alcohol research: Mediation, suppression, moderation, mediated moderation, and moderated mediation." Review of. *Alcoholism: Clinical and experimental research* 31 (s3).

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

