INCOMPLETE DRAFT FOR DISCUSSION ONLY; COMMENTS WELCOME

NOT FOR QUOTATION OR GENERAL CIRCULATION

# Causal chain analysis in systematic reviews of international development interventions

DRAFT

## 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. The types of causal relationships that can connect interventions and outcomes, are also discussed, with a focus on complex relationships and the way in which contextual factors may moderate these. We also explore the way in which specific combinations of intervention components may lead to successful interventions. We describe the process of building a logic model and the importance of this model in anchoring subsequent decisions in the systematic review process, including synthesis. 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 establishing and enhancing the salience of systematic review findings across settings, and to confirm existing theories or develop entirely new ways of understanding how interventions effect change.

## Introduction

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 influence over outcomes. This paper tracks some of this thinking, and an underlying argument that we make in this paper is that articulating 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.

## 1. Causal thinking and systematic reviews

Well-conducted systematic reviews begin with a clearly defined research question and an articulation of the conceptual framework [1]. In the context of systematic reviews of international development interventions, 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 [2]. 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 [3-5].

## 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 [4]. Reiss' review [6] identified five main accounts and perspectives through which relationships are theorised as being causal, although there is substantial overlap between these:

(i) **Counterfactual reasoning**, 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' [5], so for example 'in the absence of a microfinance intervention, there would be no improvement in poverty levels'. This form of counterfactual reasoning is partly the basis for many common forms of impact evaluation methods [7] (see also below) and is also situated in some cases within broader 'difference-making' accounts of causality [8].

(ii) **Probabilistic accounts** arise from statistical analyses of quantitative data [6] and are important to reasoning about causality in social science [5]. Many probabilistic accounts of causal relationships are based on classical linear regression models [6], 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) [9]. Studies using observational methods, for example cohort studies, also 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 underlie the interpretation of evidence from randomised controlled trials [3, 10], along with counterfactual reasoning. 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 [3, 5, 10]. However, this extrapolation can be problematic for a number of reasons [for example 3, 4, 6, 10].

(iii) **Regularity accounts** identify causal relationships through successive observation of patterns to develop regularity theories of causation [6]. 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

DRAFT

based on regularity accounts [3, 6, 11], 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 [5, 6]. 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. 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) [5]. These relationships can be highly context dependent, and the longer the causal chain, the greater the influence of context on these relationships [4]. [4].

In the case of international development interventions, failure to consider the influence of context on mechanisms can lead to unintended or harmful outcomes when interventions are transplanted from one context to another. For example, 'PlayPump', which aimed to harness children's willingness to 'play' in order to pump water through installing a merry-go-round in place of a water pump, had shown promise in some settings, but [12] when scaled up for a number of reasons [12, 13]

Among logic models (discussed below), such mechanisms may only be tentatively hypothesised at the start of the review, based more on logical reasoning than well-articulated theory, and the review process itself provides evidence for the existence and nature of the mechanism [14]. 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 [14, 15]. Developing such a mechanistic account of intervention causality draws upon evidence from other forms of causal account, e.g. counterfactual reasoning, in its creation.

(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 [6, 8]. Interventionist accounts have been criticised as being 'ideal' and not 'real', and for failing to recognise that a causal relationship between exposure and outcome may look very different from the causal relationship between intervention and a change in outcome [6, 16].

Systematic reviews of international development interventions, which are by their nature complex interventions, may draw upon several of these lenses in conceptualising and identifying causal relationships, and the evidence that we synthesise is similarly pluralistic in order to address our research questions. In fact, drawing on a number of different approaches is considered preferable because 'recalcitrant' counterexamples [17] that undermine the scope, or coverage, or validity of any one of the accounts described above when used in isolation [4, 6, 17].

## 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 [5], 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 [18, 19]. Clearly, thinking through our epistemological standpoint

DRAFT

in this way involves moving beyond the quantitative and qualitative methodological divide that has been pervasive in social science [19]. For example, systematic reviews employing quantitative synthesis methods (meta-analysis) may be drawing upon probabilistic accounts of causal relationships, but will also be drawing upon counterfactual reasoning 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. 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 [6], but it is also employed jointly alongside meta-analysis in some systematic reviews for providing mechanistic causal accounts of how effective interventions work [11, 20, 21].

Understanding our own epistemological standpoint around the types of causal accounts we are creating within our reviews 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 [3]. 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 [10], through different forms of triangulation employed during the conduct of a systematic review employing causal chain analysis [22].

For philosophers such as Illari and Russo [5], 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 [23], 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 to 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 this 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 [14, 24], 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.

## 2. Making links between inputs and outcomes

Building a causal chain involves identifying the entities (component 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' [5]. 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) – how the effect is channelled

It is this latter feature that helps to distinguish between complicated intervention and complex intervention in terms of causality [25, 26]. 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 [26]. The most simple causal relationships are those where we assume (or test) whether the intervention has 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 more complex relationships below, which also form some of the building blocks of causal chain analyses. These 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 creates the opportunities for further self-reinforcing changes [26]. 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' [27, 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.

**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 [28]. Rogers 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' [29, p70].

**Mediators, interaction effects and moderator effects**: While mediators can be represented through linear causal relationships, they are of interest as they can change their interpretation. Mediators are those factors that lie on the causal pathway between the intervention and outcome. 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 [30]. In other words, for the intervention to effect change on 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 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 these reflect study-level moderators.

**Conjuntural 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 conjuntural 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, for example an intervention component or contextual or participant characteristic. However, a single necessary characteristic may not be sufficient to trigger an outcome, and may still require the presence of other components. An example drawn from the conclusions of a literature review might be that computer/smartphone access is a necessary component of interventions that seek to enhance e-Government in sub-Saharan Africa, but is not be sufficient to trigger this outcome without a legal framework that supports implementation also being in place [31]. Enhancement of e-Government cannot occur without computer/smartphone access (access is necessary), but may not be sufficient to trigger the outcome. 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 [32] 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, as there are other routes through which the house could 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 [28, 33]. This type of 'systems thinking' is becoming increasingly common within systematic reviewing [34]. 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.



DRAFT

## 3. How can causal relationships be developed into causal chains and theories of 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 [35]. 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 [14, 24, 36]. While the use, and particularly effective and extensive use, of these techniques within systematic reviews is still at its infancy [14, 15], it is nevertheless increasingly suggested that systematic reviews include a logic model or theory of change at the outset from the protocol stage [37]. 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.

The terms 'theories of change' and 'logic models' are often used interchangeably by reviewers, largely dependent disciplinary preference [14]. Within the evaluation literature, however, a somewhat fuzzy distinction is found 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. 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 the underlying assumptions to be stated a priori [38, 39]. This distinction fits in well with the different stages of a systematic review. A logic model provides an early depiction of the components of interventions and their outcomes, but not necessarily an extensive theory-driven articulation preconditions that are needed to achieve these outcomes. New taxonomies and ways of viewing logic models are increasingly allowing for complexity into what were previously more linear forms of logic model. 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) [40]. Within the methodological literature on systematic reviewing, 'logic model' has emerged as the favoured terminology and the preferred tool for depicting intervention causal chains.

## How to develop a logic model for a systematic review as the basis for causal chain analysis

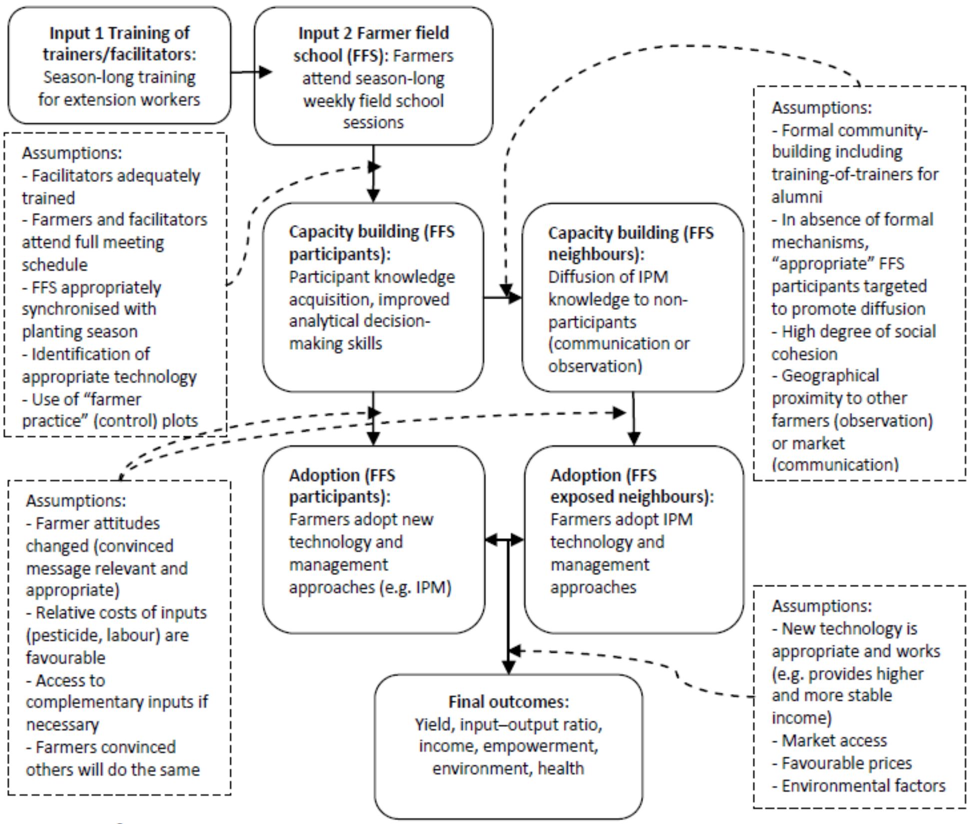
### What does a logic model look like?

Several examples of logic models exist in the systematic review literature (see [14] 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' [41, p33]. This traces the way in which outcomes (e.g. yield) are hypothesised to be determined by the presence of intermediary conditions (adoption of new technologies among participants and diffusion effects among neighbouring farmers); these are themselves shaped by a set of assumptions around moderating factors operating at a contextual level (for example market access). In turn, these adoption factors are themselves predicated on achieving a different set of circumstances, reflecting capacity issues, which are again contingent on a set of contextual factors being met [41]. While the 'type' of complex casual mechanism (e.g. any hypothesised tipping points) is not directly stated for all connections, as is rarely the case in logic models, these can sometime be expressed in footnotes to a logic model [26, 38], and the logic model itself could be used

DRAFT

as a tool to help theorise the nature of these connections [14, 42]. Furthermore, some of the complex causal relationships discussed 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 [41].

*Figure 1: Logic model for a review of farm schools (taken directly from [41])*




## Steps in building a logic model

The steps taken in developing a logic model afresh have been outlined in detail in Kneale and colleagues [14], 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 [38], as well as more recent contributions focussed on systematic reviews [24, 40, 43, 44].

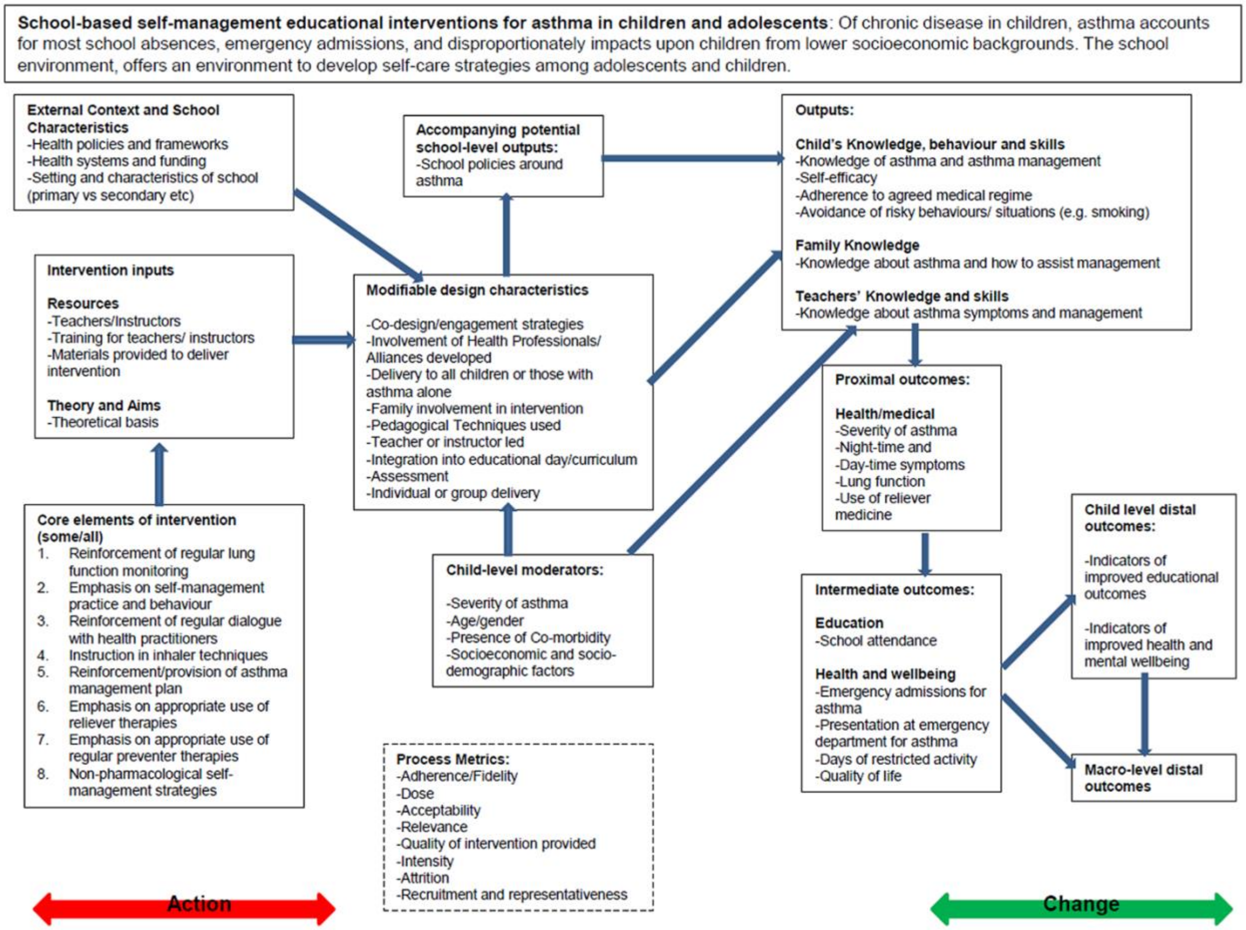
A starting point is for reviewers to familiarise themselves with the expected and intended outcomes of the intervention under study, and their potential mediating factors, as well as to consult existing logic models (or similar program theory techniques); program theory from related interventions may 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 [40]. 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 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 [4]. 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, intervention chains of intervention inputs are then specified. After completing input chains (composed of a programme's components) 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 logic model may be produced before a review team settles on a preferred model, with iterations representing 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 [45]. Some logic models may explicitly identify areas of ambiguity (e.g. 'black box' of 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 [42] for further information).

Logic models may also be used to theorise unintended outcomes and potentially negative and harmful outcomes [46]. 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 [46].

*Figure 2: Logic model for a review of farm schools (see [42])*

DRAFT




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 [47]). It is expected that the review process itself will lead to modifications in a logic model that can be used to help interpret and communicate findings [14, 48]. The final 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 [26, 38]. 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.

## Using a logic model as part of the systematic review process and in causal chain analyses

In broad terms, logic models provide a framework for 'thinking' conceptually before, during and at the end of the review [24, 36]. 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 [35], using some of the synthesis methods outlined here and elsewhere [1]. Using logic models as a framework, pathways can be systematically decomposed into lower-level pathways [35], 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.

### 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 [33]. This is 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. Similarly, for the purposes of causal chain analyses, all logic models should be process-based, and involve articulating the causal relationships between intervention components, and different mediating and target outcomes. Complex social interventions, by their nature draw upon systems theory for their identification with the expectation that complex causal relationships, including dynamic interactions with their systems of influence, take place. Building up an 'isolated description' of an intervention's causal chain [5, 49] 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 more useful for reviews of complex social interventions in International Development.

## 4. 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. Randomised controlled trials (RCTs) have been regarded as a gold standard in establishing causal relationships [50], 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 [51]). 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 [3, 50]. 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 [3, 10, 50]. 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.

DRAFT

Systematic reviews can be useful tools in helping to overcome some of these limitations. 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 'generalisable'. For example, Donaldson [52] 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). Similarly, tools exist to aid systematic reviewers to assess the underlying assumptions of RCTs, and assess the risk of bias of a trial [53], and sensitivity analyses can be employed to explore possible differential effects.

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. Reviews that are reliant on isolated descriptions of interventions will inevitably produce isolated accounts of causal relationships. This implicitly limits the generalisability of the findings, given that trial mechanisms are, at least partly, context dependent. There are also innumerable situations and reasons why conducting a RCT to evaluate the effectiveness an intervention is unfeasible, inappropriate, or unethical. Systematic reviews that have attempted synthesise evidence from RCTs, or other study designs that provide narrow 'clincher' claims, for intervention models which more frequently necessitate employing an alternative study design, have justifiably been met with criticism for a narrow scope [54]. Criteria or principles for establishing when a relationship is causal can be particularly useful, particularly for reviewers working with more diverse data, and some of these are described below.

**Tools, checklists and approaches for identifying and evaluating causal relationships**

**GRADE criteria** [55]: Although the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) criteria is used in broader terms that 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. All Cochrane reviews use GRADE to rate the quality of evidence.

**Rogers strategies** [38, 56]: 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 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** [57]: 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** [58]: 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 [59]. 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 such criteria is not necessarily helpful in identifying complex causal relationships of the type described above, such as INUS relationships for example. Similarly, many of the criteria are suitable for evaluating quantitative evidence. Frameworks for helping to explain causal relationships from qualitative data are comparatively underdeveloped [60]; this a likely reflection of the different goals of qualitative research. Based on criteria provided by Maxwell [61], the synthesis of qualitative studies providing the following forms of evidence can support causal chain analyses:

(i) comparisons within and across studies;
(ii) observation and analysis of processes and narrative connecting analysis;
(iii) understanding of discrepant cases within and across studies;
(iv) triangulation of different forms of evidence;
(v) and exploration of threats to validity.

## 5. Meta-analysis and causal chain analysis

### Traditional approaches to meta-analysis in exploring causal chains

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 [62] 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’ [63], 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 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 64, 65]. These associations are also subject to many of the same caveats of observational research, most notably

DRAFT

confounding [65, 66], 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 [67]. Nevertheless, this evidence can and is used to develop, or sometimes furnish, hypotheses of what works for whom and in what circumstances.

Extensions to these 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 [68]. 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 [11], with reviews also being undertaken that model (theory-based) complex combinations of covariates directly within meta-analysis models [42, 69].

## Extensions to 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 developmental health and wellbeing of children in low-income and middle-income countries [70]. 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' [70, p e41]. Despite their potential promise, some of the underlying assumptions of Network Meta-analysis may be difficult to substantiate, particularly for analyses that include evidence from quasi-experimental designs (although the example above did include evidence from a plurality of study designs).

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. 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 [26]. These techniques could be enhanced by the greater availability of individual level data (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 [71]). However, use of IPD meta-analyses remains scarce in the literature given the paucity of IPD data, and 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 72].



DRAFT

Perhaps one of the most direct ways of exploring strands or whole chains using quantitative synthesis, is to implement 'model-based meta-analysis' [73, 74]. 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 [74, 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's study 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 [75]. Becker presents a worked example of the stages involved [74], 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 [76]. 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 [3], but alongside other forms of synthesis described below, can be incorporated within complex and robust narratives of causal inference [4].

# 6. Alternative approaches for synthesising data on causal chains

Many different forms of synthesis can aid as part of CCA and other sources provide a detailed account of these [1, 36, 77]. We describe two of these below – QCA and its capacity to identify multiple conjunctural causation and Framework Synthesis to amalgamate different types of data – before exploring realist synthesis.

## Using framework synthesis to organise different types of evidence

Framework synthesis mirrors techniques originally used for analysing large volumes of primary qualitative data [78], but within systematic reviews has been used as a technique for amalgamating diverse data from quantitative and qualitative studies and for studying complex interventions [79]. Framework synthesis involves five key analytical stages including (i) of familiarisation with the data; (ii) theme identification (creation of a framework); (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 [80].

The causal claims resulting from techniques like framework synthesis have been aligned with hypothetico-deductive reasoning [3, 5], where the aim is to uncover enough, sufficiently varied, and novel evidence to substantiate the hypothesis if it were true [3] (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 that can provide a casual clinch [3]. For CCA it presents a more holistic option in marshalling different forms of evidence to populate different causal strands, and through the

DRAFT

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 [79, 80]. 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 [21]. A framework, developed from a previous review of community engagement [81], 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 restricted to 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) [82]. This review started through 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' [82, p5/27], including case reports and case series. Although the description of the methods omitted explicit naming of the processes of Framework Synthesis, the description provided appeared to encompass several stages. 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.

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. In addition, when focussed upon similar strands of the causal chain, it upholds other principles in causal attribution, principally triangulation. It is also closely related to other techniques, particularly the use of logic models [83]. 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.

## 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 [84]. 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 [11]. QCA is being used within systematic reviews both to further understand the results of meta-

DRAFT

analyses [for example 21], to develop theories to test within meta-analyses [for example 42], 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 [84]. 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; (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 [85]. 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 [86]; identifying such sufficient relationships is usually the more common purpose and outcome of QCA. When used in combination with meta-analysis, QCA emerges as a powerful technique of understanding how the organization of intervention components can cause changes in outcomes.

## 7. How are realist approaches used in reviews of International Development interventions?

Unlike systematic reviews examining 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 key to generate desired changes [87]. 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 leads 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' [87]. 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 [88, 89].

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' [89, p452]. As outlined by Pawson [87], building on similar causal mechanisms operating under different contexts

DRAFT

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 1 for further details and examples from international development [90-93]).

DRAFT

*Table 1: Examples of Realist Reviews of International Development Interventions*

| Study | Intervention focus | Initial theoretical framework | Types of evidence included | Rigour and transparency | Process of identifying and configuring C-M-O |
|---|---|---|---|---|---|
| **Dieleman et al 2009 [90]** | Human resources management (HRM) interventions<br>Seven types of interventions in scope and classified according to the three HRM-intervention levers | Developed a framework to facilitate understanding of mechanisms which shows that there are variety of interested mechanisms<br>Included studies that did not report on the underlying assumptions of how the interventions should bring about to change. | All types of studies | Bias in the evaluation studies | 'We systematically assessed outcome, context, and mechanisms through which the intervention produced its outcomes.' [90, p2]<br>Mechanisms were identified if they were reported by the study authors. 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.<br>Considered theories of behaviour change<br>Limited reporting on the context, implementation, mechanism, underlying assumptions of how the intervention should bring about change. |
| **Kane et al 2010 [91]** | Use of community health workers (CHW) | Not stated | RCTs | Not reported | Mechanisms were included only when they were either researched or discussed by the authors of the RCTs [91, P4]. [90, P4].<br>Iterative and discussion between review teams, a common understanding of C-M-O was arrived<br>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:<br>• an anticipation of being valued by the community,<br>• a perception of improvement in social status, and having a valuable social role<br>• a sense of relatedness with and accountability to the beneficiaries" |
| **Westhorp et al 2014 [92]** | Community accountability | 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 (actors whose decision-making has been changed, the reasoning that underlies the changed decision, and |

DRAFT

| | | | | | |
|---|---|---|---|---|---|
| | | | | | outcome of the different decision, and refined the initial programme theory<br>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 [93]** | School accountability systems: assessment, monitoring and inspection | 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<br>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):<br>- teacher-level individual incentives,<br>- pressures from school leadership and external stakeholders for results, or<br>- teachers' recognition that the incentive is of value and merits additional effort. |

DRAFT

Identifying and developing the theoretical framework at the onset of the review process is useful in defining scope of the review and identifying generic causal mechanisms before synthesising evidence [77, 90, 92, 93]. For example, a 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, the framework was developed to facilitate understanding of mechanisms of human resource management interventions to improve availability, productivity, responsiveness, and competency of worker's performance in low and middle-income countries [90].

Quality appraising in systematic reviews aims to evaluate whether the methods employed are appropriate and the findings are reliable [1]. Whilst realist synthesis considers 'rigour', it also recognises quality 'an emergent property' [93, p22] throughout the process of review [94]. In addition, relevance is considered by extent to which the findings support or refute the initial theoretical framework [93, 95]. 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 [91]. For example, reviewers typically report several round of reading and rereading 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 [91-93]. 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.

## Conclusions: Causal chain analysis in systematic reviews of international development interventions

Taking a CCA approach enables reviewer 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 [96]. 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, 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 review and RCT. 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" [96, p273].

DRAFT

In fact, synthesis of a broad range of types of qualitative evidence has flourished over recent decades [1, 97] with new approaches continually developed [11, 98]. 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 causal reasoning in order to evaluate different strands of the causal chain. 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 [4, 6, 17]. This is similar to some of the ideas advocated within mixed studies/mixed methods reviews, although Causal Chain Analyses might be focussed less around the integration of qualitative and quantitative data which may scaffold the same 'link' [99], and is focussed more heavily 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.

## 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 [1] for an overview). The following represent loose principles that could be applied in the conduct of future CCA for International Development 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; the involvement of intervention stakeholders in the development of the logic model; the representation of potential harms (dark logic [46]); the representation of contextual factors; and the extensive use of the logic model within different review processes.
3. **Development of research questions that relate to hypothesised causal relationships:** 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 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 causal longer and more complex 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 made clear 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 [100, p83]. This also showed which pathways were not assessed during the review process.

## 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 provide a conclusive mechanistic account of how and how much an intervention changes an outcome. Incorporating different data may be one strategy to overcome this limitation [see also 101], 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 [68, 69] and the more extensive use of other meta-analytic frameworks such as Network Meta-Analysis. Some of these approaches are also related to the use of QCA, which was identified as 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 [102]. Without an understanding of how the intervention works, such evidence of cost-effectiveness becomes the type of evidence that provides the 'clincher' [3], 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



DRAFT

that improves the external validity of findings and identifies where further investigation is most needed' [100].

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. However, the logic model is a guiding hypothesis, often based on poor or incomplete descriptions of interventions [103], and despite incorporating the elements of good practice described above, may oversimplify (and thus incorrectly specify) a complex systems-based intervention [26]. 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' [26, 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.

## References


1. Gough D, Oliver S, Thomas J: **An introduction to systematic reviews**: Sage; 2017.
2. Rogers PJ: **Causal models in program theory evaluation**. *New directions for evaluation* 2000, **2000**(87):47-55.
3. Cartwright N: **Hunting causes and using them: Approaches in philosophy and economics**: Cambridge University Press; 2007.
4. Krieger N, Davey Smith G: **The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology**. *International journal of epidemiology* 2016, **45**(6):1787-1808.
5. Illari P, Russo F: **Causality: Philosophical theory meets scientific practice**: OUP Oxford; 2014.
6. Reiss J: **Causation in the social sciences: Evidence, inference, and purpose**. *Philosophy of the Social Sciences* 2009, **39**(1):20-40.
7. Gertler PJ, Martinez S, Premand P, Rawlings LB, Vermeersch CM: **Impact evaluation in practice**. Washington DC: World Bank Publications; 2016.
8. Vandenbroucke JP, Broadbent A, Pearce N: **Causality and causal inference in epidemiology: the need for a pluralistic approach**. *International journal of epidemiology* 2016, **45**(6):1776-1786.
9. Wooldridge JM: **Introductory econometrics: A modern approach**: Nelson Education; 2015.
10. Cartwright N: **What are randomised controlled trials good for?** *Philosophical studies* 2010, **147**(1):59.
11. Thomas J, O'Mara-Eves A, Brunton G: **Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example**. *Systematic reviews* 2014, **3**(1):1-14.
12. Ika LA, Donnelly J: **Success conditions for international development capacity building projects**. *International Journal of Project Management* 2017, **35**(1):44-63.
13. UNICEF: **An Evaluation of the PlayPump water system as an appropriate Technology for Water, Sanitation and Hygiene Programmes**. In.; 2007.
14. Kneale D, Thomas J, Harris K: **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* 2015.


DRAFT

15. Maden M, Cunliffe A, McMahon N, Booth A, Carey GM, Paisley S, Dickson R, Gabbay M: **Use of programme theory to understand the differential effects of interventions across socio-economic groups in systematic reviews—a systematic methodology review**. *Systematic reviews* 2017, **6**(1):266.
16. Kelly MP, Russo F: **Causal narratives in public health: the difference between mechanisms of aetiology and mechanisms of prevention in non-communicable diseases**. *Sociology of health & illness* 2017.
17. Reiss J: **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* 2012, **43**(4):769-777.
18. Gough D, Thomas J: **Commonality and diversity in reviews**. In: *An Introduction to Systematic Reviews.* edn. Edited by Gough D, Oliver S, Thomas J. London: Sage; 2017.
19. Gough D, Thomas J, Oliver S: **Clarifying differences between review designs and methods**. *Systematic reviews* 2012, **1**(1):1.
20. Chandler J, Thomas J, Sutcliffe K, Kahwati L, Kneale D: **Applying current philosophical insights on causality using Qualitative Comparative Analysis as an additional synthesis in systematic reviews to address complex interventions**. In: *Gobal Evidence Summit.* Cape Town, South Africa; 2017.
21. Brunton G, Caird J, Kneale D, Thomas J, Richardson M: **Community engagement for health via coalitions, collaborations and partnerships: a systematic review and meta-analysis**. In. London: EPPI-Centre, UCL Institute of Education; 2015.
22. Hales D: **An introduction to triangulation**. *UNAIDS monitoring and evaluation* 2010.
23. Snilstveit B: **Systematic reviews: from 'bare bones' reviews to policy relevance**. *Journal of development effectiveness* 2012, **4**(3):388-408.
24. Anderson LM, Petticrew M, Rehfuess E, Armstrong R, Ueffing E, Baker P, Francis D, Tugwell P: **Using logic models to capture complexity in systematic reviews**. *Research synthesis methods* 2011, **2**(1):33-42.
25. Lewin S, Hendry M, Chandler J, Oxman AD, Michie S, Shepperd S, Reeves BC, Tugwell P, Hannes K, Rehfuess EA: **Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT_SR)**. *BMC medical research methodology* 2017, **17**(1):76.
26. Rogers PJ: **Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions**. *Evaluation* 2008, **14**(1):29-48.
27. Lund C, De Silva M, Plagerson S, Cooper S, Chisholm D, Das J, Knapp M, Patel V: **Poverty and mental disorders: breaking the cycle in low-income and middle-income countries**. *The lancet* 2011, **378**(9801):1502-1514.
28. Shiell A, Hawe P, Gold L: **Complex interventions or complex systems? Implications for health economic evaluation**. *BMJ: British Medical Journal* 2008, **336**(7656):1281.
29. Vaessen J, Rivas A, Duvendack M, Palmer-Jones R, Leeuw FL, van Gils G, Lukach R, Holvoet N, Bastiaensen J, Hombrados JG: **The effect of microcredit on women's control over household spending in developing countries: a systematic review**. *Campbell Systematic Reviews* 2014, **10**(8).
30. Vollmer S, Khan S, Le Thi Ngoc Tu A, Pasha SS: **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**. In. Oslo, Norway: Campbell Collaboration; 2017.
31. Nkohkwo QN-a, Islam MS: **Challenges to the Successful Implementation of e-Government Initiatives in Sub-Saharan Africa: A Literature Review**. *Electronic Journal of e-Government* 2013, **11**(1).

DRAFT

32. Mackie JL: **Causes and conditions**. *American philosophical quarterly* 1965, **2**(4):245-264.
33. Clark AM: **What are the components of complex interventions in healthcare? Theorizing approaches to parts, powers and the whole intervention**. *Social Science & Medicine* 2013, **93**:185-193.
34. Noyes J, Gough D, Lewin S, Mayhew A, Michie S, Pantoja T, Petticrew M, Pottie K, Rehfuess E, Shemilt I: **A research and development agenda for systematic reviews that ask complex questions about complex interventions**. *Journal of clinical epidemiology* 2013, **66**(11):1262-1270.
35. Wirtz PW: **Advances in causal chain development and testing in alcohol research: Mediation, suppression, moderation, mediated moderation, and moderated mediation**. *Alcoholism: Clinical and Experimental Research* 2007, **31**(s3).
36. Waddington H, White H, Snilstveit B, Hombrados JG, Vojtkova M, Davies P, Bhavsar A, Eyers J, Koehlmoos TP, Petticrew M: **How to do a good systematic review of effects in international development: a tool kit**. *Journal of development effectiveness* 2012, **4**(3):359-387.
37. Collaboration C: **Protocol Template**. In.; 2017.
38. Funnell SC, Rogers PJ: **Purposeful program theory: effective use of theories of change and logic models**, vol. 31. San Francisco, CA: John Wiley & Sons; 2011.
39. Clark H, Anderson AA: **Theories of Change and Logic Models: Telling Them Apart**. In: *American Evaluation Association.* Atlanta, Georgia; 2004.
40. Rohwer A, Pfadenhauer L, Burns J, Brereton L, Gerhardus A, Booth A, Oortwijn W, Rehfuess E: **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* 2017, **83**:37-47.
41. Waddington H, Snilstveit B, Hombrados JG, Vojtkova M, Anderson J, White H: **Farmer field schools for improving farming practices and farmer outcomes in low-and middle-income countries: a systematic review**. *Campbell systematic reviews* 2014, **10**(6).
42. Harris KM, Kneale D, Lasserson TJ, McDonald VM, Grigg J, Thomas J: **School-based self management interventions for asthma in children and adolescents: a mixed methods systematic review**. *The Cochrane Library* 2015.
43. Baxter SK, Blank L, Woods HB, Payne N, Melanie R, Elizabeth G: **Using logic model methods in systematic review synthesis: describing complex pathways in referral management interventions**. *BMC Medical Research Methodology* 2014, **14**(1):62.
44. Pfadenhauer L, Rohwer A, Burns J, Booth A, Lysdahl KB, Hofmann B, Gerhardus A, Mozygemba K, Tummers M, Wahlster P: **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**. In.; 2016.
45. Rees R, Oliver S: **Stakeholder perspectives and participation in reviews**. In: *An Introduction to Systematic Reviews.* edn. Edited by Gough D, Oliver S, Thomas J. London: Sage Publications; 2012.
46. Bonell C, Jamal F, Melendez-Torres GJ, Cummins S: **"Dark logic": theorising the harmful consequences of public health interventions**. *Journal of epidemiology and community health* 2014, **69**(1):95-98.
47. Booth A, Carroll C: **Systematic searching for theory to inform systematic reviews: is it feasible? Is it desirable?** *Health Information & Libraries Journal* 2015, **32**(3):220-235.
48. Rehfuess EA, Booth A, Brereton L, Burns J, Gerhardus A, Mozygemba K, Oortwijn W, Pfadenhauer LM, Tummers M, Wilt GJ: **Towards a taxonomy of logic models in systematic reviews and health technology assessments: a priori, staged and iterative approaches**. *Research synthesis methods* 2017.

DRAFT

49. Craver CF: **Role functions, mechanisms, and hierarchy**. *Philosophy of science* 2001, **68**(1):53-74.
50. Cartwright N: **Are RCTs the gold standard?** *BioSocieties* 2007, **2**(1):11-20.
51. Petticrew M, Roberts H: **Evidence, hierarchies, and typologies: horses for courses**. *Journal of Epidemiology & Community Health* 2003, **57**(7):527-529.
52. Donaldson SI, Street G, Sussman S, Tobler N: **Using meta-analyses to improve the design of interventions**. In: *Handbook of program development for health behavior research and practice.* edn. Edited by Sussman S. Thousand Oaks, California: Sage; 2001: 449-466.
53. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savović J, Schulz KF, Weeks L, Sterne JA: **The Cochrane Collaboration's tool for assessing risk of bias in randomised trials**. *Bmj* 2011, **343**:d5928.
54. Petticrew M, Shemilt I, Lorenc T, Marteau T, Melendez-Torres G, O'Mara-Eves A, Stautz K, Thomas J: **Alcohol advertising and public health: systems perspectives versus narrow perspectives**. *Journal of epidemiology and community health* 2016:1-5.
55. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ: **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* 2008, **336**(7650):924.
56. Rogers P: **Overview: Strategies for Causal Attribution: Impact Evaluation No. 6**. In: *Methodological Briefs.* Florence, Italy: UNICEF; 2014.
57. Howick J, Glasziou P, Aronson JK: **The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation'contribute?** *Journal of the Royal Society of Medicine* 2009, **102**(5):186-194.
58. Hill AB: **The environment and disease: association or causation?** *Journal of the Royal Society of Medicine* 2015, **108**(1):32-37.
59. Weed DL: **Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related**. *International Journal of Epidemiology* 2000, **29**(3):387-390.
60. Rychetnik L, Frommer M, Hawe P, Shiell A: **Criteria for evaluating evidence on public health interventions**. *Journal of epidemiology and community health* 2002, **56**(2):119-127.
61. Maxwell JA: **Using qualitative methods for causal explanation**. *Field methods* 2004, **16**(3):243-264.
62. Mekasha TJ, Tarp F: **Aid and growth: What meta-analysis reveals**. *The journal of development studies* 2013, **49**(4):564-583.
63. Duvendack M, Hombrados JG, Palmer-Jones R, Waddington H: **Assessing 'what works' in international development: meta-analysis for sophisticated dummies**. *Journal of development effectiveness* 2012, **4**(3):456-471.
64. Petticrew M, Tugwell P, Kristjansson E, Oliver S, Ueffing E, Welch V: **Damned if you do, damned if you don't: subgroup analysis and equity**. *Journal of epidemiology and community health* 2011.
65. Thompson SG, Higgins JP: **Can meta-analysis help target interventions at individuals most likely to benefit?** *The Lancet* 2005, **365**(9456):341-346.
66. Sun X, Briel M, Walter SD, Guyatt GH: **Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses**. *Bmj* 2010, **340**:c117.
67. Kneale D, O'Mara-Eves A, Thomas J: **It's a mean world for generalisability: Knowledge claims of generalisability from meta-analyses of public health interventions**. under review.
68. Kneale D, Thomas J, O'Mara-Eves A, Wiggins RD: **Enhancing the portability of public health intervention review evidence for localised decision-making** under review.

DRAFT

69. Harris KM, Kneale D, Lasserson TJ, McDonald VM, Grigg J, Thomas J: **School-based self management interventions for asthma in children and adolescents: a mixed methods systematic review**. *The Cochrane Library* forthcoming(4).
70. Welch VA, Ghogomu E, Hossain A, Awasthi S, Bhutta ZA, Cumberbatch C, Fletcher R, McGowan J, Krishnaratne S, Kristjansson E: **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* 2017, **5**(1):e40-e50.
71. Riley RD, Lambert PC, Abo-Zaid G: **Meta-analysis of individual participant data: rationale, conduct, and reporting**. *Bmj* 2010, **340**.
72. van Eijk AM, Hill J, Larsen DA, Webster J, Steketee RW, Eisele TP, ter Kuile FO: **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* 2013, **13**(12):1029-1042.
73. Becker BJ: **Examining theoretical models through research synthesis: The benefits of model-driven meta-analysis**. *Evaluation & the health professions* 2001, **24**(2):190-217.
74. Becker BJ: **Model-based meta-analysis**. In: *The handbook of research synthesis and meta-analysis. Volume 2*, edn. Edited by Cooper H, Hedges LV, Valentine JC; 2009: 377-395.
75. Whiteside MF, Becker BJ: **Parental factors and the young child's postdivorce adjustment: a meta-analysis with implications for parenting arrangements**. In.: American Psychological Association; 2000.
76. Skrondal A, Rabe-Hesketh S: **Structural equation modeling: categorical variables**: Wiley Online Library; 2005.
77. Snilstveit B, Oliver S, Vojtkova M: **Narrative approaches to systematic review and synthesis of evidence for international development policy and practice**. *Journal of development effectiveness* 2012, **4**(3):409-429.
78. Ritchie J, Spencer L: **Qualitative data analysis for applied policy research**. *The qualitative researcher's companion* 2002, **573**(2002):305-329.
79. Brunton G, Oliver S, Thomas J: **Applying framework synthesis to understand complexity in systematic reviews: A systematic review of methods.** In: *23rd Cochrane Colloquium.* Vienna, Austria; 2015.
80. Booth A, Carroll C: **How to build up the actionable knowledge base: the role of 'best fit'framework synthesis for studies of improvement in healthcare**. *BMJ Qual Saf* 2015, **24**(11):700-708.
81. O'Mara-Eves A, Brunton G, McDaid G, Oliver S, Kavanagh J, Jamal F, Matosevic T, Harden A, Thomas J: **Community engagement to reduce inequalities in health: a systematic review, meta-analysis and economic analysis**. *Public Health Research* 2013, **1**(4).
82. Krauer F, Riesen M, Reveiz L, Oladapo OT, Martínez-Vega R, Porgo TV, Haefliger A, Broutet NJ, Low N, Group WZCW: **Zika virus infection as a cause of congenital brain abnormalities and Guillain–Barré syndrome: systematic review**. *PLoS medicine* 2017, **14**(1):e1002203.
83. Harden A, Thomas J, Cargo M, Harris J, Pantoja T, Flemming K, Booth A, Garside R, Hannes K, Noyes J: **Cochrane Qualitative and Implementation Methods Group Guidance Paper 4: Methods for integrating qualitative and implementation evidence within intervention effectiveness reviews**. *Journal of clinical epidemiology* 2017.
84. Ragin CC: **Redesigning social inquiry: Fuzzy sets and beyond**, vol. 240: Wiley Online Library; 2008.
85. Langer L, Tripney J, Erasmus Y, Tannous N, Chisoro C, Opondo M, Zigana L, Obuku E, van Rooyen C, Stewart R: **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**

DRAFT

**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; in press.
86. Sutcliffe K, Richardson M, Rees R, Melendez-Torres G, Stansfield C, Thomas J: **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* 2016.
87. Pawson R: **Evidence-based policy: The promise of 'realist synthesis'**. *Evaluation* 2002, **8**(3):340-358.
88. Pawson R, Tilley N: **Realistic evaluation**: Sage; 1997.
89. Mallett R, Hagen-Zanker J, Slater R, Duvendack M: **The benefits and challenges of using systematic reviews in international development research**. *Journal of development effectiveness* 2012, **4**(3):445-455.
90. Dieleman M, Gerretsen B, van der Wilt GJ: **Human resource management interventions to improve health workers' performance in low and middle income countries: a realist review**. *Health Research Policy and Systems* 2009, **7**(1):7.
91. Kane SS, Gerretsen B, Scherpbier R, Dal Poz M, Dieleman M: **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* 2010, **10**(1):286.
92. Westhorp G, Walker B, Rogers P, Overbeeke N, Ball D, Brice G: **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* 2014.
93. Eddy-Spicer D, Ehren M, Bangpan M, Khatwa M, Perrone F: **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; 2016.
94. Pawson R: **Evidence-based policy: a realist perspective**: Sage; 2006.
95. Greenhalgh J: **Realist synthesis**. In: *Studying Organizations Using Critical Realism: A Practical Guide.* edn. Edited by Edwards P, O'Mahoney J, Vincent S. Oxford: Oxford University Press; 2014.
96. Cornish F: **Evidence synthesis in international development: a critique of systematic reviews and a pragmatist alternative**. *Anthropology & medicine* 2015, **22**(3):263-277.
97. Thomas J, Harden A: **Methods for the thematic synthesis of qualitative research in systematic reviews**. *BMC medical research methodology* 2008, **8**(1):45.
98. Sutcliffe K, Thomas J, Stokes G, Hinds K, Bangpan M: **Intervention Component Analysis (ICA): a pragmatic approach for identifying the critical features of complex interventions**. *Systematic reviews* 2015, **4**(1):1.
99. Pluye P, Hong QN: **Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews**. *Annual review of public health* 2014, **35**.
100. Willey B, Smith Paintain L, Mangham L, Car J, Armstrong Schellenberg J: **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**. In. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London 2013.
101. White H: **Theory-based impact evaluation: principles and practice**. *Journal of development effectiveness* 2009, **1**(3):271-284.
102. White H: **Current challenges in impact evaluation**. *The European Journal of Development Research* 2014, **26**(1):18-30.



DRAFT

103. Hoffmann TC, Oxman AD, Ioannidis JP, Moher D, Lasserson TJ, Tovey DI, Stein K, Sutcliffe K, Ravaud P, Altman DG: **Enhancing the usability of systematic reviews by improving the consideration and description of interventions**. *bmj* 2017, **358**:j2998.

DRAFT