Synthesizing evidence on complex interventions: how meta-analytical, qualitative, and mixed-method approaches can contribute

Mark Petticrew, Eva Rehfuess, Jane Noyes, Julian P.T. Higgins, Alain Mayhew, Tomas Pantoja, Ian Shemilt, Amanda Sowden

Abstract

Objectives: Although there is increasing interest in the evaluation of complex interventions, there is little guidance on how evidence from complex interventions may be reviewed and synthesized, and the relevance of the plethora of evidence synthesis methods to complexity is unclear. This article aims to explore how different meta-analytical approaches can be used to examine aspects of complexity; describe the contribution of various narrative, tabular, and graphical approaches to synthesis; and give an overview of the potential choice of selected qualitative and mixed-method evidence synthesis approaches.

Study Design and Setting: The methodological discussions presented here build on a 2-day workshop held in Montebello, Canada, in January 2012, involving methodological experts from the Campbell and Cochrane Collaborations and from other international review centers (Anderson L, Petticrew M, Chandler J, et al. Introduction: systematic reviews of complex interventions. In press). These systematic review methodologists discussed the broad range of existing methods and considered the relevance of these methods to reviews of complex interventions.

Results: The evidence from primary studies of complex interventions may be qualitative or quantitative. There is a wide range of methodological options for reviewing and presenting this evidence. Specific contributions of statistical approaches include the use of meta-analysis, meta-regression, and Bayesian methods, whereas narrative summary approaches provide valuable precursors or alternatives to these. Qualitative and mixed-method approaches include thematic synthesis, framework synthesis, and realist synthesis. A suitable combination of these approaches allows synthesis of evidence for understanding complex interventions.

Conclusion: Reviewers need to consider which aspects of complex interventions should be a focus of their review and what types of quantitative and/or qualitative studies they will be including, and this will inform their choice of review methods. These may range from standard meta-analysis through to more complex mixed-method synthesis and synthesis approaches that incorporate theory and/or user’s perspectives.

Keywords: Systematic review; Complex interventions; Evidence synthesis; Evaluation; Mixed methods; Meta-analysis

1. Introduction: why is guidance needed on synthesizing evidence on complex interventions?

Definitions of complex interventions emphasize the need to take account of a range of dimensions of complexity [2]. Typical definitions include interactions and synergies between intervention components, interactions between the intervention and characteristics of the context in which it is implemented, and the need to understand the impacts of other mediators and moderators on the effects of the intervention [3]. Complex interventions may also be
characterized by phase changes (where there may be long periods when little appears to be happening, then large changes occur) and feedback loops (where the outcomes of the intervention influence the intervention itself [4,5]). Table 1 lists various sources of complexity that have been identified in the academic literature.

This article aims to:

- explore how different meta-analytical approaches can be used to examine aspects of complexity;
- describe the contribution of various narrative, tabular, and graphical approaches to synthesis, and how they can be used to explore complex interventions; and
- give an overview of the potential choice of selected qualitative and mixed-method evidence synthesis approaches

It discusses approaches that are of most value for addressing questions of effectiveness and methods for synthesizing qualitative evidence.

The goal of a systematic review, whether it incorporates meta-analysis, is often seen as simply “pooling” (aggregating) evidence to achieve an overall quantitative or narrative summary, usually about whether an intervention works or does not. However, such an approach may be just the first step in a longer process of exploration, not just of whether an intervention works, but how it works, in what populations/subpopulations, and in what circumstances and contexts. Here, the primary goal of synthesis is not to simply integrate the studies, and thus risk obscuring differences, but to understand commonalities and differences between studies. Disaggregation may therefore be as important as aggregation. The aim is therefore not simply combination but integration of evidence of different types from different sources in such a way that it sheds new light on different aspects of complexity. However, the preliminary stage of a complex review may still involve answering the most straightforward question about effectiveness. For example, a preliminary stage of some synthesis approaches to examine the effects of complex interventions (e.g., meta-regression) involves conducting a standard meta-analysis.

There are many potential sources of complexity in the relationship between an intervention and its effects. Where these sources are very influential, they should be investigated using appropriate methods. Fig. 1 introduces different options for specific synthesis methods, broadly characterized as:

- Quantitative meta-analytical approaches, involving the statistical synthesis of data from primary quantitative studies, and narrative summary approaches (by which we mean summarizing study findings in narrative, tabular, or graphical form, describing differences in context, validity, and other study characteristics);
- Qualitative synthesis that summarizes the results of qualitative studies without aiming to generate theory;
- Qualitative evidence synthesis approaches, whereby a variety of different methods can be selected to develop a theory of why an intervention works (or not), explore the experiences of beneficiaries or other relevant stakeholders and examine implementation factors—the product of which can stand alone or can be integrated with the quantitative synthesis; and
- Mixed-method synthesis approaches whereby quantitative, mixed-method, and qualitative evidence is integrated in a single review [6,7].

Although Bayesian meta-analysis appears under “mixed-method synthesis” in Fig. 1, it is debatable whether it is truly “mixed methods” as the qualitative data are transformed into quantitative data. Also, “narrative summary” appears under the “quantitative synthesis” although it could also be seen as a separate option (i.e., one that narratively summarizes quantitative data but does not result in a pooled effect size, unlike meta-analysis).

2. Synthesizing evidence on complex interventions

The task of synthesizing evidence on complex interventions can be daunting. However it can be made simpler by breaking it down into a series of stages (Fig. 2). First, we might consider the potential sources of complexity in the relationship between an intervention and health or other
Table 1. Mapping the main sources of complexity onto study design inclusion criteria and type of synthesis

<table>
<thead>
<tr>
<th>Source of complexity</th>
<th>Study design or source of data for this type of complexity</th>
<th>Synthesis may involve or be helped by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple components. Synergies/interactions (between components of an intervention). Dysynergistic interactions are also possible, where the effect of the intervention will be less than the summation of effects of the individual actions [30], for example, palliative care comprises a range of distinct and complementary services, including specific symptom management (i.e., pain, breathlessness), physiotherapy, as well as psychosocial and spiritual support [31]. Connectivity—where individual components of an intervention are linked together in a system, so they influence each other [32]. Feedback loops and other nonlinear relationships (cannot be arranged along a simple input—output line, such as where changes in behavior create the conditions for behavior to change further—for example, uptake in cycling results in more cyclists, which means that cycling becomes the norm, encouraging more people to take up cycling; Galea et al. [15]). Self-organization, adaptivity, and evolution over time [32,38]. Phase transitions</td>
<td>Evidence of the independent effects of components of the intervention, and interactions between those components, may be available in the form of either quantitative or qualitative data. Studies with multiple arms, particularly those with factorial designs, may explore these effects. Individual studies with different configurations of components may be included in a review allowing indirect comparisons between studies [33–35]. Qualitative studies carried out alongside trials or may be described in qualitative or quantitative research (e.g., structural equation modeling). Longitudinal studies carried out as part of process evaluations may be valuable [32]. As aforementioned, may be described in qualitative or quantitative research and may require longitudinal studies carried out as part of process evaluations [32]. Longitudinal data, for example, changes in direction or size of effects over time may be observed in studies with multiple data points (e.g., Interrupted Time Series studies). Qualitative data may also be available to describe phase changes. Data on multiple health and nonhealth outcomes may come from any type of evaluative study; qualitative studies may also show the range and nature (direction) of effects. Any type of evaluative study; for example, cluster RCTs may provide outcome data at both cluster, and individual level; studies may collect data from individuals about effects not just on themselves but on their families, communities, and so forth. External data sources (e.g.,</td>
<td>Investigating heterogeneity at the study level using subgroup analysis, meta-regression, or network meta-analysis or at the participant level with individual participant data meta-analysis (possibly using hierarchical models). Synthesis of qualitative and mixed-method evidence on key stakeholder perspectives may explain heterogeneity, and implementation evidence can help explain differences in interactions due to health system factors and so forth. Investigating heterogeneity at the study level using subgroup analysis or meta-regression or at the participant level with individual participant data meta-analysis (possibly using hierarchical models). Exploring stakeholder views, which may help in explaining synergies, interactions, and connectivity. Incorporating expert judgments about likelihood of feedback loops using bias adjustment. Exploring stakeholder views which may help in explaining relationships and communication to support behavior change and spread of innovation. Identifying a range of sources on context, and on how the intervention has changed over time; including nonacademic sources (e.g., policy documents and stakeholder experiences [10]). Investigating heterogeneity at the study level using subgroup analysis or meta-regression, or at the participant level with individual participant data meta-analysis (possibly using hierarchical models). Modeling correlated outcomes using multivariate meta-analysis. Exploring personal accounts of patients, which may contribute to understanding of the multiple proximal and distal outcomes and their perspectives on the importance placed on specific outcomes. Repeating the synthesis with a different unit of analysis (e.g., in cluster randomized trials, the analysis can focus on cluster-level outcomes or participant-level outcomes); undertaking individual participant data meta-analysis;</td>
</tr>
<tr>
<td>Multiple outcomes [3,4,39], for example, improved stove or cleaner fuel interventions to reduce household air pollution in developing country homes result in multiple health (e.g., acute and chronic respiratory diseases, cardiovascular disease, birth weight) and nonhealth benefits (e.g., fuel savings, time savings) [40,41]. Effects at different levels [3,38], for example, impact of vaccination programs at individual and population levels through herd immunity.</td>
<td>Process evaluations, studies describing implementation, and policy documents and other sources [3,10,38].</td>
<td></td>
</tr>
</tbody>
</table>
outcomes; then, we can frame appropriate research questions and hypotheses relating to these sources of complexity. This may be aided by the development of a logic model [9] that can clarify the reviewer’s reasoning by enhancing understanding of the theory of change underpinning programs or policies and by setting out the initial assumptions about the causal pathways between intervention, mechanisms, and outcomes. Such models also provide a framework for identifying study inclusion/exclusion criteria, guiding the search strategy (such as search terms and databases), identifying relevant outcomes, and examining differences among studies and along dimensions of interest. The search for relevant evidence will follow from this (e.g., qualitative or quantitative studies, or both, may be relevant, as well as other types of evidence such as empirical case studies or policy documents); data extraction forms will then need to be developed to capture data on the relevant aspects of complexity. Finally, we can conduct the synthesis of the various elements. As for any systematic review, the final synthesis needs to take account of study quality and the confidence that we can have in the findings of individual studies.

To illustrate this process, we may decide (perhaps on the basis of existing theory or knowledge of the intervention) that complexity derives from many different interacting components. We may also decide that although phase changes are relevant, they are not a key feature of interest. In this case, we may focus on locating primary studies of different designs for evidence of interactions between components of the intervention. Qualitative studies, for example, may describe how different components are expected to work together, and quantitative studies may suggest that the intervention is more effective when more than one component is used. Unfortunately, in practice, clear quantitative evidence of such interactions may be difficult to find because the relevant analyses are seldom conducted as part of primary studies.

Interactions between intervention and context are commonly believed to be important in complex interventions. By context, we mean the particular geographical, cultural, and social environment and the organizational and political systems in which an intervention or program takes place. Information about context may come from a wide range of sources. The first step in the synthesis may involve describing the range of contexts within which the intervention has been implemented, and extracting relevant data that suggest how the context influences mechanisms and outcomes.

This is not straightforward. Systematic reviews of complex interventions typically provide details of the intervention (e.g., components, duration, and who delivered it), the setting (e.g., where and when the intervention took place), and the population receiving the intervention. However, less attention is provided by the review to information about the context in which the intervention is implemented. Without information about the local policy and wider political environment, economic factors, and information relating to the physical or structural environment, including
resources available, it is not possible to assess how intervention effects might be modified by any one of, or a combination of, these factors. This information is also critical for assessment of applicability and transferability of the intervention.

Contextual information can sometimes be obtained from the evaluation studies themselves as well as from process evaluations and qualitative studies, but other sources of evidence such as policy documents may also be valuable [10], particularly when the intervention being evaluated is a policy. For example, a review examining the impact of presumed consent legislation on organ donation rates concluded that it was not possible to explore in detail the likely impact of the legislation change [11]. This was because the evaluation studies did not fully report or investigate the influence of contextual factors such as how the legislation was interpreted (level of consultation with the deceased’s family), awareness-raising programs, media coverage of transplantation issues, and any infrastructure changes. Additional analysis of policy and other historical documents providing contextual information relating to each country introducing presumed consent legislation may have provided the missing data. The value of subject experts—including users of the review—should not be overlooked. It is now good practice to ensure that review questions are defined together with potential users of the review, and this is equally important in reviews focused on complex interventions. Users or stakeholders may include funders, practitioners, policy and other decision makers, and those who are the subjects of the interventions, such as patients and members of the public. Such stakeholders may help with the identification and interpretation of contextual influences on interventions (and other aspects of the review, including statistical and other aspects, including Bayesian approaches, and the development of logic models).

3. How might complexity relate to approaches to synthesis?

An earlier article in this series was concerned with how to frame the question for a review of complex interventions

---

**Fig. 1.** Synthesis strategy and synthesis method options for complex intervention reviews—adapted with permission from Noyes and Lewin [8]. The final box (product) on the left hand side reflects the scenario whereby a team undertake a review of effectiveness and another undertake a qualitative review—but to explore complexity and explain heterogeneity and effect, an additional synthesis or integration of the two review products needs to happen in a subsequent step. Note also that although Bayesian synthesis is frequently described as mixed methods, it may also be seen as solely quantitative.
We reiterate here that the choice of empirical approach, study design, and component data to be synthesized and the choice of a meta-analytical, narrative, qualitative, or mixed-method approach depend critically on the research question. Investigation of complexity in a systematic review is not an end in itself but should be strongly driven by the research question(s). In some situations, investigating aspects of complexity may be valuable, but in others it will not. It is not necessary for every systematic review to be concerned with feedback loops, phase changes, interactions between interventions components, and so on, unless this is important to understand the intervention and/or is helpful to users. Moreover, although the technical solutions promised by evolving systematic review methods will be useful for understanding complex interventions, making full use of the potential of systematic reviews in relation to complex interventions also requires broader systems thinking. This may be helped by the use of logic models as aforementioned or may be developed through the application of complex system approaches and systems epidemiology.

Evidence on different aspects of complexity (Table 1) derives from many different types of data. This has implications for each step in a typical systematic review, including framing the question(s), translating review questions into study eligibility criteria, searching for eligible studies, synthesizing the results of included studies, and examining heterogeneity. For example, evidence on phase changes may derive from quantitative studies, in which data are collected over long periods of time (e.g., interrupted time series), or qualitative studies. The overall synthesis strategy (a single method or mix of approaches, as per Fig. 1) will therefore need to consider this fact, and appropriate specific synthesis methods should be chosen to reflect this.

The distinction between “aggregative” and “configurative” reviews may also be useful here. The former tend to use quantitative prespecified methods to test theory and answer well-defined questions using empirical observations (a deductive approach). Reviews who take the latter approach, however, tend to ask more open-ended questions and use more iterative methods to generate and explore theory (an inductive method). Either or both classes of approach (evidence synthesis strategies) may be used in the case of systematic reviews that include a focus on issues of complexity. Research questions about the components of the intervention and interactions between them, for example, may be addressed using an aggregative approach; these may lend themselves to tightly specified hypotheses, which can be tested with quantitative data from the primary studies. A more configurative approach may draw additionally on qualitative and other evidences to explore aspects of complexity.

Systematic review authors are also aiming to produce more rounded and nuanced understandings of the effects of such interventions. Different methodological frameworks for integrating quantitative and qualitative research to synthesize evidence for the effects of complex interventions and their mechanisms of action are described in an earlier article in this series. As such, integrating diverse forms of evidence often requires the parallel conduct and integration of two or more different types of synthesis (Fig. 1). Chapter 20 in the Cochrane handbook and additional supplemental guidance produced by the Cochrane Qualitative and Implementation methods group provide guidance on approaches to integration of the products of reviews. Whether these syntheses are undertaken within the framework of a single review or as two or more separate parallel reviews is in part an operational issue and in part a methodological issue. In some circumstances, it may be judged appropriate to conduct a series of parallel reviews (rather than attempt to incorporate all data into the same review), and then integrate the findings at the end. For example, a systematic review (and/or meta-analysis) may be conducted to summarize the evidence on effectiveness from intervention studies, with evidence on process and implementation issues then being synthesized in a separate systematic review. Indeed, this is what was done in a Cochrane review of school feeding interventions. The effectiveness review used meta-analysis to summarize the evidence on differential effectiveness and explore effect modifiers; whereas principles from realist review were subsequently used to analyze process data.

In the following sections, we briefly describe the respective (and often complementary) roles and contributions of
meta-analytical, narrative, and qualitative/mixed-method approaches to synthesize evidence for understanding complex interventions.

4. The contribution of statistical approaches

The traditional strengths of applying meta-analytical synthesis methods in Cochrane and other systematic reviews of the effects of interventions are that they can address questions about: (1) whether an overall effect exists across a larger body of evidence than an individual study; (2) whether effects are consistent across studies; (3) what is the actual magnitude and variation of effects across studies; and (4) whether particular study-level factors are associated with the magnitude of effect. These questions are equally applicable to systematic reviews of complex interventions.

Table 2 provides a general description of the most important meta-analytical approaches (i.e., subgroup analyses, meta-regression, multivariate meta-analysis, network meta-analysis, and individual participant data meta-analysis) and broader statistical approaches (i.e., hierarchical models, Bayesian methods, bias adjustment, and causal diagram—based analysis) and illustrates their specific relevance to complex interventions. Much can be achieved with standard meta-analytical methods, for example, using random-effects models to recognize the usually large clinical/study heterogeneity and subgroup analyses to explore the effects of components of the intervention. Careful a priori specification of subgroup analyses or meta-regression analyses is necessary to reduce the likelihood of spurious findings, and as usual post hoc analyses should be clearly identified [12]. Subsets of studies with particular characteristics (such as particular combinations of intervention components) can be identified and their findings compared with other studies and with the overall average effect.

Comparisons can be made between individual studies to shed light on interactions between intervention components, although such analyses will suffer from low power and be susceptible to confounding because of other differences between the studies [13]. It may also be helpful to distinguish between the influence of study-level characteristics on outcomes and the influence of intervention/comparator-level characteristics on outcomes as the meta-analytical techniques that are applicable will differ. For example, univariate meta-regression analysis requires data on study-level characteristics, and if intervention or comparison-level characteristics cannot be summarized at this level, hierarchical models or multivariate approaches may be required. The effect of context on outcomes can be explored in a similar manner.

Given the sophistication of the analyses that may be required and the probable need for a range of secondary hypotheses relating to aspects of complexity, a clear and detailed protocol is needed, setting out the analyses and their relationship to the research questions. Indeed, a detailed protocol is probably even more important in reviews of more complex interventions than in reviews of simpler interventions.

5. The contribution of narrative summary approaches

Narrative summary methods can be used in conjunction with meta-analytical and other quantitative approaches to explain and summarize reasons for differences in findings between studies. This systematic organization and presentation of the data can help the reviewer and reader to identify themes across studies and can facilitate the testing of prespecified theory by exploring similarities and differences among studies [22]. This is an important first step in any synthesis, whether or not a meta-analysis is performed. Systematically structured tables and graphical presentations of findings are important aspects of the synthesis and can offer an analytical approach to explore aspects of complexity. Tables provide details of study populations, methods, and findings and can organize and present information on the context and setting within which each study was conducted. Likewise, they can be organized and stratified to show how effects vary by population, setting, context, or any other study characteristic.

Graphical summaries too can be used as an analytical tool to help explore sources of variability among studies. Harvest plots, for example, have been used to present findings stratified by population and intervention and to show findings across multiple outcomes and settings. They could also be used to stratify by other aspects of complexity, for example, by showing how study results vary with different approaches to implementation [1]. Fig. 3 illustrates findings from studies of the effects of tobacco control interventions. It includes information on study design, outcomes, and direction of effect. This format could be adapted to incorporate other study or intervention characteristics, including those relating to complexity.

Heterogeneity is considered in another article in this series [23], but a few additional comments are relevant here. It is well known that statistical attempts to understand causes of heterogeneity have minimal power when there are very few studies, such that any associations discovered are likely to be spurious. The same considerations apply to the use of nonstatistical approaches to understand causes of heterogeneity, so there is a danger of overinterpreting any patterns identified. The risk of spurious conclusions can be reduced by having clearly specified research questions that draw on relevant theory. Logic models again have a key role to play here, as does qualitative evidence [9].

6. The contribution of qualitative and mixed-method approaches

As shown previously and in Table 2, evidence synthesis conducted in the tradition of systematic reviews of
## Table 2. Statistical approaches to evidence synthesis with relevance to complex interventions

<table>
<thead>
<tr>
<th>General description</th>
<th>Relevance to complex interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Specific meta-analytical approaches</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup analysis</strong> splits the studies or the participants according to population, intervention, or contextual characteristics and examines differences in effect estimates across these subgroups. Characteristics are chosen as those likely to have an impact on the size or direction of the intervention effect [44].</td>
<td></td>
</tr>
<tr>
<td><strong>Application:</strong></td>
<td>Compare subgroups of participants exposed to different intervention components.</td>
</tr>
<tr>
<td></td>
<td>The impact of major contextual influences (e.g., different implementation mechanisms) can be explored.</td>
</tr>
<tr>
<td><strong>Challenges:</strong></td>
<td>Differences between subgroups may not be because of the characteristic used to define subgroups (in other words, there may be confounding because subgroup comparisons are observational by nature) [45].</td>
</tr>
<tr>
<td></td>
<td>Sufficient number of studies are required to provide convincing results (ideally at least five studies for each subgroup); systematic reviews are rarely able to assess more than one or two factors at the same time.</td>
</tr>
<tr>
<td></td>
<td>Performing multiple subgroup analyses may involve multiple statistical tests, with inflated risk of spurious findings [44].</td>
</tr>
<tr>
<td></td>
<td>Characteristics may be correlated, hampering interpretation.</td>
</tr>
<tr>
<td></td>
<td>Standard subgroup analyses cannot generally be used to compare time points within studies.</td>
</tr>
<tr>
<td><strong>Meta-regression</strong> explores the relationship across studies between study characteristics and effect sizes, offering a generalization of subgroup analysis. It draws on the same principles as regression analysis in primary studies, allowing for the effects of continuous and/or categorical variables to be modeled but is conducted at the level of studies rather than at the level of study participants. In addition to testing for statistical significance, the amount of between-study variation in effect sizes that can be explained by the characteristic(s) can be quantified using an index analogous to the $R^2$ index in regression analysis of primary data [44,46].</td>
<td></td>
</tr>
<tr>
<td><strong>Application:</strong></td>
<td>Explore sources of heterogeneity in effect sizes and their relative importance, for example, in relation to intervention components, degree of tailoring, and various contextual influences.</td>
</tr>
<tr>
<td></td>
<td>Examine phase transitions by modeling impact of study duration.</td>
</tr>
<tr>
<td><strong>Challenges:</strong></td>
<td>Relationships may not be due to the characteristics included in the meta-regression model (in other words, there may be confounding because meta-regression analyses are observational by nature) [44].</td>
</tr>
<tr>
<td></td>
<td>Sufficient numbers of studies are required to provide convincing results (ideally at least 10 studies for each characteristic included in the model) [19,44].</td>
</tr>
<tr>
<td></td>
<td>Characteristics may be correlated, hampering interpretation.</td>
</tr>
<tr>
<td><strong>Multivariate meta-analysis</strong> allows each study to contribute two or more (possibly correlated) effect estimates to the meta-analysis. For example, these may be effects on two outcomes, effects at two-time points, or effect sizes for different interventions included in the same study [47].</td>
<td></td>
</tr>
<tr>
<td><strong>Application:</strong></td>
<td>Multivariate meta-analysis can facilitate a joint analysis of intermediate outcomes with downstream outcomes for the same participants, which allows for the correlation in the treatment effects to be estimated.</td>
</tr>
<tr>
<td></td>
<td>Multivariate meta-analysis of two or more different time points may be used to explore phase transitions.</td>
</tr>
<tr>
<td><strong>Challenges:</strong></td>
<td>Requires specialist statistical expertise and software.</td>
</tr>
<tr>
<td></td>
<td>Need some information about within-study correlations.</td>
</tr>
<tr>
<td><strong>Network meta-analysis</strong> compares multiple interventions simultaneously by analyzing studies making different comparisons in the same analysis. It is a complex form of meta-regression, and if some studies have multiple intervention groups, then it is a multivariate meta-regression. Different components of a complex intervention may be treated as different interventions, and assumptions made about whether the effects components are additive or interact with one another [34].</td>
<td></td>
</tr>
<tr>
<td><strong>Application:</strong></td>
<td>Investigation of whether interventions with a particular component (or combination of components) are more effective.</td>
</tr>
<tr>
<td><strong>Challenges:</strong></td>
<td>Requires specialist statistical expertise and software.</td>
</tr>
</tbody>
</table>

(Continued)
### General description

#### Individual participant data meta-analysis

Draws on the original research data for each study participant in each included study. Individual participant data meta-analysis is usually conducted in two stages. In the first stage, data in individual studies are reanalyzed in a consistent way. In the second stage, the results of each individual study are combined in a summary estimate of effect, analogous to standard meta-analysis. Alternatively, one-stage methods are available and are an application of hierarchical models (see following) [19].

#### Broader statistical approaches

**Hierarchical models** are based on the fact that participants are nested within studies that, in turn, are nested within the meta-analysis. Most standard meta-analysis models are hierarchical models, but the idea can be extended to the specific nature of the studies at hand, to account for clustering at various levels. Variability is apportioned to different levels of the hierarchy, for example, in a meta-analysis of cluster-randomized trials, we might have a between-participant component, a (within-study) between-cluster component, and a between-study component. Characteristics of each type of unit can also be modeled using regression approaches [48].

**Bayesian methods** follow a different philosophy of statistics from the classical frequentist statistics. Insights gained from new data (i.e., the studies included in a systematic review) are combined with prior knowledge, following the idea of updating knowledge with evidence. In practical terms, prior knowledge is incorporated in meta-analysis by specifying a prior distribution to describe uncertainty in the effect size and/or the likely extent of between-study variation [49].

**Bias-adjustment** may be used to combine effect estimates that are subject to multiple and different sources of bias. This is an application of Bayesian methods. Prior knowledge or belief about the biases may derive from empirical studies of bias [50] or opinions elicited formally from experts [51,52]. The priors have the effect of adjusting effect estimates for individual studies. These adjusted estimates may then be combined in a standard random-effects meta-analysis.

**Causal diagram-based analysis** draws on the principles of DAGs as a formal tool for identifying covariates that need to be included in a model for the effect estimated to be unbiased. Although, to date, causal diagrams have only been applied at the level of primary studies, the DAG principles can be extended to evaluate studies for their inclusion in meta-analysis. Causal diagrams may be seen as a formal statistical extension of logic models that make analytical assumptions explicit [9]. A major challenge is the upfront specification of one or more causal diagrams based on existing knowledge supplemented with biologically plausible assumptions [53].

### Relevance to complex interventions

- Assumes that all interventions included in the “network” are equally applicable to all populations and contexts of the studies included.

**Application:**
- Could overcome problem of many sources of heterogeneity in studies of complex interventions, allowing the analysis to focus on actual differences in intervention type and delivery and potential interactions between interventions and participant characteristics.
- Very useful for examining outcomes at multiple levels, for example, by conducting analyses for aggregated outcomes.
- Where outcomes in primary studies were assessed at multiple points in time would facilitate examination of phase transitions.

**Challenges:**
- Rarely feasible for complex interventions, given large heterogeneity in primary studies.
- Time consuming

**Abbreviation:** DAGs, directed acyclic graphs.
effectiveness can contribute to address a variety of questions of relevance to complex interventions. In many circumstances, however, answering these questions will require a broader approach to enquiry [18]. Many methods to synthesize qualitative studies have been described in the literature, with some of them having been applied to issues of complexity, whereas others have potential application but few published examples [24]. Consequently, for those wishing to embark on a systematic review of broader questions, the choice of method is far from straightforward.

Table 3 presents some of the most commonly used qualitative and mixed-method approaches, describes whether these are mainly aggregative (i.e., testing theory, generating a mostly utilitarian outcome) or configuring/interpretive (i.e., generating or exploring theory, generating a mostly theoretical outcome) in nature, and highlights their potential application to complex interventions [17,25].

Beyond a mostly aggregative vs. mostly configuring aim of a systematic review, the suitability of different methods across a range of empirical approaches (e.g., qualitative, quantitative, theoretical, empirical case studies, policy documents) can also help with the categorization of different methods. Additional criteria have been described by Barnett-Page and Thomas [24] in relation to where the methods sit on a spectrum between (1) epistemology of idealisms vs. epistemology of realism and (2) no iteration vs. extensive iteration. In methods that embrace an idealistic

stance, the assumption is that all knowledge is constructed (e.g., critical interpretative synthesis), whereas the epistemology of realism assumes that the world is seen as it really is (e.g., framework synthesis). Approaches with a limited degree of iteration (e.g., thematic synthesis) pursue a highly structured approach to select, organize, and tabulate the primary research data, whereas revisions of initial decisions and processes as the systematic review progresses are common practice in approaches with a high degree of iteration (e.g., critical interpretative synthesis) [24].

Ultimately, the choice of method will be influenced by the nature of the research in relation to the aforementioned criteria and practical considerations, such as the existing skill set and experience of the research team and resource requirements in terms of personnel, time, and funding. With many of these approaches, an expert advisory group may help with conceptual clarity and decision making. It is also important to note that purely qualitative synthesis approaches may need an additional phase, involving integrating the results with an effectiveness review.

7. Conclusions and areas for future research

Despite the range of methods already available, there are many areas where new methodological research is needed. Among these are further development of methods for incorporating diverse sources of evidence in single reviews, technical examination of specific analytical methods [26,27], and further assessment of the role of Bayesian analyses to combine quantitative evidence with prior distributions derived from qualitative evidence.

In many reviews of complex interventions, it makes little sense to capture the impact of an intervention on a single primary outcome, therefore the means of dealing with multiple health and nonhealth outcomes are in need of development. The role of surrogate and intermediate outcomes warrants further exploration. For example, are room concentrations of pollutants a valid surrogate for a variety of short-term (e.g., low birth weight, child pneumonia) and long-term (e.g., cancer, cardiovascular disease) outcomes? A particular unresolved problem is how best to synthesize evidence from outcomes measured at multiple levels such as at the level of the individual, the community, and the population. For example, how do we synthesize evidence on individual vs. herd immunity in the case of immunization programs, or on individual and neighborhood effects in the case of urban regeneration programs? [28] Identifying and dealing with interactions between components is a key aspect of evidence synthesis in this area, yet existing methods are not sensitive. Developing approaches to separate “whole-package” intervention effects vs. the effects of “active ingredient” and/or combination of both would therefore be useful. Such interactions are often mentioned in the literature but methods for dealing with them rarely described.
<table>
<thead>
<tr>
<th>General description</th>
<th>Relevance to complex interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Qualitative approaches</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Meta-ethnography</strong> serves to (1) translate the findings of different primary research studies on the same phenomenon into each other to generate overarching themes, concepts or metaphors (reciprocal translational analysis); (2) identify and explain contradictions and differences that exist between the various studies (refutational synthesis); and (3) develop a picture of the whole phenomenon under study from studies of its parts (line-of-argument synthesis) [56]. The approach is demonstrated in the review by Campbell et al. [57] on lay experiences of diabetes and diabetes.</td>
<td></td>
</tr>
<tr>
<td><strong>Application:</strong></td>
<td>Primarily generates theory—such as program theory, implementation theory, or an explanatory theory of why the intervention works or not, hypotheses for future testing or comparison with trial outcomes.</td>
</tr>
<tr>
<td><strong>Characteristics:</strong></td>
<td>Involves putting together written interpretive accounts to produce a higher-order synthesis [24].</td>
</tr>
<tr>
<td><strong>Mixed-method approaches</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Critical interpretative synthesis</strong> represents an adaptation of meta-ethnography as well as grounded theory in the synthesis of both qualitative and quantitative evidence—such as the review by Dixon-Woods et al. [60] on access to healthcare by vulnerable groups. This method aims to generate a synthesizing argument or theory and entails a highly iterative approach to refining the research question and obtaining the primary research sample, as well as data analysis. Consequently, critical interpretative synthesis applies a method of assessing quality of primary research studies according to their contribution to theory development rather than methodological attributes [61].</td>
<td></td>
</tr>
<tr>
<td><strong>Application:</strong></td>
<td>Primarily generates theory such as program theory, implementation theory, or an explanatory theory of why the intervention works or not, hypotheses for future testing or comparison with trial outcomes.</td>
</tr>
<tr>
<td><strong>Characteristics:</strong></td>
<td>Requires significant methodological skill and experience—therefore can be resource intensive and may take time to engage with the evidence and develop theory.</td>
</tr>
</tbody>
</table>

**Challenges:**
- Requires further interpretation by policy makers and practitioners
- Requires significant methodological skill and experience with qualitative methods
- Therefore can be resource-intensive and may take time to engage with the evidence and develop theory.
- Not inherently reproducible or transparent because of highly iterative nature of interpretive process.

**Thematic synthesis (various traditions)** uses thematic analysis techniques to identify themes across primary research studies. The synthesis component entails an iterative process of inductively grouping themes into overarching categories that capture the similarities, differences, and relationships between the themes for the purpose of generating hypotheses about the phenomenon under study. Broadly, this process comprises two main steps, that is, line-by-line coding of the relevant parts of individual studies and the iterative generation of categories [58]. This synthesis approach has been described by Harden in al. [59] in a review of studies of people’s views.

| **Application:** | Suitable for providing complementary information about implementation, acceptability, or other features where an effectiveness review is conducted in parallel. |
| **Characteristics:** | Draws on critical realistic epistemology [23] |
| **Mixed-method approaches** | |
| **Critical interpretative synthesis** represents an adaptation of meta-ethnography as well as grounded theory in the synthesis of both qualitative and quantitative evidence—such as the review by Dixon-Woods et al. [60] on access to healthcare by vulnerable groups. This method aims to generate a synthesizing argument or theory and entails a highly iterative approach to refining the research question and obtaining the primary research sample, as well as data analysis. Consequently, critical interpretative synthesis applies a method of assessing quality of primary research studies according to their contribution to theory development rather than methodological attributes [61]. | |
| **Application:** | Primarily generates theory such as program theory, implementation theory, or an explanatory theory of why the intervention works or not, hypotheses for future testing or comparison with trial outcomes. |
| **Characteristics:** | Requires further interpretation by policy makers and practitioners
- Requires significant methodological skill and experience—therefore can be resource intensive and may take time to engage with the evidence and develop theory. |

**Challenges:**
- Requires further interpretation by policy makers and practitioners
- Requires significant methodological skill and experience—therefore can be resource intensive and may take time to engage with the evidence and develop theory. |

**Not inherently reproducible or transparent because of highly iterative nature of interpretive process.**
Framework synthesis is based on a method previously developed by Ritchie and Spencer [62]. It constitutes a deductive approach that uses an a priori coding framework that can be informed by theory, reflect the review questions, and derived from the researchers’ discussions and a preliminary reading of included studies. Framework synthesis produces a map of each key dimension identified in the synthesis and the nature of the influence and association with other dimensions; see, for example, the review by Oliver et al. [63] which they used to help develop a framework for public involvement in health services research [64].

Realist synthesis
A “realist” review aims to determine what works for whom, in what circumstances, and why/how mechanisms work or not (Pawson et al. [65])—for example, Pawson’s review of “naming and shaming.” A program theory (or theories) is made explicit about how an intervention is meant to work and its anticipated impacts. Reviewers then seek out supporting or conflicting evidence in a highly iterative way [66].

Narrative synthesis is a mixed-method synthesis approach that starts with developing a preliminary theory of why the intervention and or implementation works. Includes a toolbox for transforming and translating qualitative and quantitative evidence [67]. This approach is illustrated in a review of the implementation of child injury prevention initiatives [68].

Application:
- Primarily explores theory—a priori framework may be informed by theory, product may explain why theory works or not. May be specifically helpful with interrogating relationships. Nonetheless, some framework synthesis products are sufficiently rich to develop conceptual models and new theoretical insights/hypotheses.

Characteristics:
- More critical realistic epistemology
- Limited iteration
- Transformation of primary studies
- Directly informs policy makers and practitioners
- Highly transparent approach

Challenges:
- Challenging to execute by an inexperienced team
- Requires significant methodological skill and experience—therefore can be resource-intensive and may take time to engage with the evidence and develop theory.
- Not inherently reproducible or transparent because of highly iterative nature of interpretive process.
- Some realist syntheses include additional primary research processes such as focus groups with recipients of care to develop or confirm theory.

Application:
- Most applicable for complex interventions where evidence of effect may be lacking.
- Generates and explains theory such as program theory, implementation theory, or an explanatory theory of why the intervention works or not, hypotheses for future testing, mechanisms, or comparison with trial outcomes.

Characteristics:
- More critical realistic epistemology
- High level iteration
- Transformation and translation of primary studies
- Directly informs policy makers and practitioners

Challenges:
- Requires significant methodological skill and experience—therefore can be resource-intensive and may take time to engage with the evidence and develop theory.
- Not inherently reproducible or transparent because of highly iterative nature of interpretive process.

Adapted and expanded from Barnett-Page and Thomas [24] and Hannes and Lockwood [25].
There is also a methodological agenda for qualitative and mixed-method reviews. Like other reviews, these vary in quality, and this may be because of lack of equivalent publication guidelines to CONSORT and PRISMA, although guidelines for meta-narrative and realist reviews have recently been published [29]. Although qualitative and mixed-method approaches offer potential opportunities to explore sources of complexity at several different junctures in the review process, there are as yet few high-quality published examples to show what these methods could contribute to understand the effects of complex interventions. Some methods—such as critical interpretive synthesis—have thus far only been used by the originators and few others. There is also little guidance on how best to structure separate quantitative and qualitative review products to aid investigation of aspects of complexity. Most methods lack specification, and review authors have fairly quickly added their own modifications and adaptations, which has concerned some originators [29]. Importantly, the quality of the review depends on the quality and experience of the review team, especially with interpretive reviews that inevitably rely to varying degrees on the ability of the authors to interpret evidence and develop new theory in a careful and appropriate way. More experience and better guidance are required in using theory at various stages of the review process to interpret evidence and in particular to help understand complexity. In particular, we require greater understanding about the importance of proximity of qualitative evidence to the original intervention, and whether generic downstream evidence about an implemented intervention can be considered in the same way as specific evidence collected concurrently alongside RCTs. This is an emerging and rapidly developing field, and it is likely that reviewers will take up and adapt many existing methods, including some not covered in this article such as summarizing effect sizes using the “median of medians” approach sometimes used in EPOC (Cochrane Effective Practice and Organisation of Care group) reviews. Some of these adaptations may even challenge existing systematic review practices. Systematic review methods are flexible but it is also important that the impact of changes to existing review approaches are evaluated and the findings of those evaluations reported fully—what worked and what did not work—in the appropriate methodological journals.

References