

# Synthesising quantitative evidence in systematic reviews of complex health interventions

Julian P T Higgins,<sup>1</sup> José A López-López,<sup>1</sup> Betsy J Becker,<sup>2</sup> Sarah R Davies,<sup>1</sup> Sarah Dawson,<sup>1</sup> Jeremy M Grimshaw,<sup>3,4</sup> Luke A McGuinness,<sup>1</sup> Theresa H M Moore,<sup>1,5</sup> Eva A Rehfuess,<sup>6</sup> James Thomas,<sup>7</sup> Deborah M Caldwell<sup>1</sup>

**To cite:** Higgins JPT, López-López JA, Becker BJ, et al. Synthesising quantitative evidence in systematic reviews of complex health interventions. *BMJ Glob Health* 2019;4:e000858. doi:10.1136/bmjgh-2018-000858

**Handling editor** Soumyadeep Bhaumik

Received 29 March 2018  
Revised 13 August 2018  
Accepted 14 August 2018

## ABSTRACT

Public health and health service interventions are typically complex: they are multifaceted, with impacts at multiple levels and on multiple stakeholders. Systematic reviews evaluating the effects of complex health interventions can be challenging to conduct. This paper is part of a special series of papers considering these challenges particularly in the context of WHO guideline development. We outline established and innovative methods for synthesising quantitative evidence within a systematic review of a complex intervention, including considerations of the complexity of the system into which the intervention is introduced. We describe methods in three broad areas: non-quantitative approaches, including tabulation, narrative and graphical approaches; standard meta-analysis methods, including meta-regression to investigate study-level moderators of effect; and advanced synthesis methods, in which models allow exploration of intervention components, investigation of both moderators and mediators, examination of mechanisms, and exploration of complexities of the system. We offer guidance on the choice of approach that might be taken by people collating evidence in support of guideline development, and emphasise that the appropriate methods will depend on the purpose of the synthesis, the similarity of the studies included in the review, the level of detail available from the studies, the nature of the results reported in the studies, the expertise of the synthesis team and the resources available.

## BACKGROUND

Public health and health service interventions are typically complex. They are usually multifaceted, with impacts at multiple levels and on multiple stakeholders. Also, the systems within which they are implemented may change and adapt to enhance or dampen their impact.<sup>1</sup> Quantitative syntheses ('meta-analyses') of studies of complex interventions seek to integrate quantitative findings across multiple studies to achieve a coherent message greater than the sum of their parts. Interest is growing on how the standard systematic review and

## Summary box

- ▶ Quantitative syntheses of studies on the effects of complex health interventions face high diversity across studies and limitations in the data available.
- ▶ Statistical and non-statistical approaches are available for tackling intervention complexity in a synthesis of quantitative data in the context of a systematic review.
- ▶ Appropriate methods will depend on the purpose of the synthesis, the number and similarity of studies included in the review, the level of detail available from the studies, the nature of the results reported in the studies, the expertise of the synthesis team and the resources available.
- ▶ We offer considerations for selecting methods for synthesis of quantitative data to address important types of questions about the effects of complex interventions.

meta-analysis toolkit can be enhanced to address complexity of interventions and their impact.<sup>2</sup> A recent report from the Agency for Healthcare Research and Quality and a series of papers in the *Journal of Clinical Epidemiology* provide useful background on some of the challenges.<sup>3-6</sup>

This paper is part of a series to explore the implications of complexity for systematic reviews and guideline development, commissioned by WHO.<sup>7</sup> Clearly, and as covered by other papers in this series, guideline development encompasses the consideration of many different aspects,<sup>8</sup> such as intervention effectiveness, economic considerations, acceptability<sup>9</sup> or certainty of evidence,<sup>10</sup> and requires the integration of different types of quantitative as well as qualitative evidence.<sup>11 12</sup> This paper is specifically concerned with methods available for the synthesis of quantitative results in the context of a systematic review on the effects of a complex intervention. We aim to point those collating evidence in support



©World Health Organization 2019. Licensee BMJ.

For numbered affiliations see end of article.

### Correspondence to

Professor Julian P T Higgins;  
julian.higgins@bristol.ac.uk

of guideline development to methodological approaches that will help them integrate the quantitative evidence they identify. A summary of how these methods link to many of the types of complexity encountered is provided in table 1, based on the examples provided in a table from an earlier paper in the series.<sup>1</sup> An annotated list of the methods we cover is provided in table 2.

We begin by reiterating the importance of starting with meaningful research questions and an awareness of the purpose of the synthesis and any relevant background knowledge. An important issue in systematic reviews of complex interventions is that data available for synthesis are often extremely limited, due to small numbers of relevant studies and limitations in how these studies are conducted and their results are reported. Furthermore, it is uncommon for two studies to evaluate exactly the same intervention, in part because of the interventions' inherent complexity. Thus, each study may be designed to provide information on a unique context or a novel intervention approach. Outcomes may be measured in different ways and at different time points. We therefore discuss possible approaches when data are highly limited or highly heterogeneous, including the use of graphical approaches to present very basic summary results. We then discuss statistical approaches for combining results and for understanding the implications of various kinds of complexity.

In several places we draw on an example of a review undertaken to inform a recent WHO guideline on protecting, promoting and supporting breast feeding.<sup>13</sup> The review seeks to determine the effects of interventions to promote breast feeding delivered in five types of settings (health services, home, community, workplace, policy context or a combination of settings).<sup>8</sup> The included interventions were predominantly multicomponent, and were implemented in complex systems across multiple contexts. The review included 195 studies, including many from low-income and middle-income countries, and concluded that interventions should be delivered in a combination of settings to achieve high breastfeeding rates.

## THE IMPORTANCE OF THE RESEARCH QUESTION

The starting point in any synthesis of quantitative evidence is a clear purpose. The input of stakeholders is critical to ensure that questions are framed appropriately, addressing issues important to those commissioning, delivering and affected by the intervention. Detailed discussion of the development of research questions is provided in an earlier paper in the series,<sup>1</sup> and a subsequent paper explains the importance of taking context into account.<sup>9</sup> The first of these papers describes two possible perspectives. A *complex interventions perspective* emphasises the complexities involved in conceptualising, specifying and implementing the intervention per se, including the array of possibly interacting components and the behaviours required to implement

it. A *complex systems perspective* emphasises the complexity of the systems into which the intervention is introduced, including possible interactions between the intervention and the system, interactions between individuals within the system and how the whole system responds to the intervention.

The simplest purpose of a systematic review is to determine whether a particular type of complex intervention (or class of interventions) is effective compared with a 'usual practice' alternative. The familiar PICO framework is helpful for framing the review:<sup>14</sup> in the PICO framework, a broad research question about effectiveness is uniquely specified by describing the participants ('P', including the setting and prevailing conditions) to which the intervention is to be applied; the intervention ('I') and comparator ('C') of interest, and the outcomes ('O', including their time course) that might be impacted by the intervention. In the breastfeeding review, the primary synthesis approach was to combine all available studies, irrespective of setting, and perform separate meta-analyses for different outcomes.<sup>15</sup>

More useful than a review that asks 'does a complex intervention work?' is one that determines the situations in which a complex intervention has a larger or smaller effect. Indeed, research questions targeted by syntheses in the presence of complexity often dissect one or more of the PICO elements to explore how intervention effects vary both within and across studies (ie, treating the PICO elements as 'moderators'). For instance, analyses may explore variation across participants, settings and prevailing conditions (including context); or across interventions (including different intervention components that may be present or absent in different studies); or across outcomes (including different outcome measures, at different levels of the system and at different time points) on which effects of the intervention occur. In addition, there may be interest in how aspects of the underlying system or the intervention itself mediate the effects, or in the role of intermediate outcomes on the pathway from intervention to impact.<sup>16</sup> In the breastfeeding review, interest moved from the overall effects across interventions to investigations of how effects varied by such factors as intervention delivery setting, high-income versus low-income country, and urban versus rural setting.<sup>15</sup>

## The role of logic models to inform a synthesis

An earlier paper describes the benefits of using system-based logic models to characterise a priori theories about how the system operates.<sup>1</sup> These provide a useful starting point for most syntheses since they encourage consideration of all aspects of complexity in relation to the intervention or the system (or both). They can help identify important mediators and moderators, and inform decisions about what aspects of the intervention and system need to be addressed in the synthesis. As an example, a protocol for a review of the health effects of environmental interventions to reduce the consumption

**Table 1** Quantitative synthesis possibilities to address aspects of complexity

Aspect of complexity of interest	Examples of potential research question(s)	Synthesis possibilities	Further discussion
What 'is' the system? How can it be described?	What are the main influences on the health problem? How are they created and maintained? How do these influences interconnect?	Map the system, defining pathways and influences. Draw a logic model based on the key aspects for the research question at hand as a basis for thinking about the quantitative synthesis.	See companion paper, <sup>1</sup> and section 2.1.
Interactions between components of complex interventions	What is the independent and combined effect of the individual components? How do the components work along and in combination to produce effects? (How do they interact to produce outcomes?)	Consider methods such as meta-regression, network meta-analysis and component-based approach that address intervention components, using models that allow investigation of interactions among components.	See sections 5.2 and 6.
Interactions of interventions with context and adaptation	Do the effects of the intervention appear to be context-dependent?	Consider subgroup analysis and meta-regression to examine how features of context impact on effect sizes.	See section 5.2.
System adaptivity (how does the system change?)	(How) does the system change when the intervention is introduced?	Identify behaviours or actions that might be affected, and consider these as outcomes in meta-analysis or meta-regression analyses. To account for correlations among them, multivariate methods might be considered.	See section 8.
Emergent properties	Which aspects of the system are affected? Does this potentiate or dampen its effects?	Identify units (eg, individuals or organisations) whose behaviour or actions might be affected, and consider these as outcomes in meta-analysis or meta-regression. Multilevel models might be appropriate to capture the different levels of impact, although may require access to individual participant data.	See sections 5.2 and 8.
Non-linearity and phase changes	What are the effects (anticipated and unanticipated) which follow from this system change?	Identify other possible effects of the intervention, and consider these as outcomes in meta-analysis or meta-regression analyses. Consider model-driven meta-analysis or mathematical models (including simulation approaches) to investigate these further.	See section 8, <a href="#">box 2</a> and <a href="#">box 3</a> .
Positive (reinforcing) and negative (balancing) feedback loops	How do effects change over time? What explains change in the effectiveness of the intervention over time?	Identify important time points and address these in separate meta-analyses, or using meta-regression analyses. Consider mathematical models to predict how effects might change over time.	See sections 5 and 8, and <a href="#">box 3</a> .
	Are the effects of an intervention dampened/suppressed by other aspects of the system (eg, contextual influences)?	Consider model-driven meta-analysis or mathematical models to investigate these. Consider subgroup analysis and meta-regression to examine how features of the system impact on effect sizes.	See sections 7 and 8, <a href="#">boxes 2 and 3</a> . See section 5.2.

Continued

Table 1 Continued

Aspect of complexity of interest	Examples of potential research question(s)	Synthesis possibilities	Further discussion
Multiple (health and non-health) outcomes	What changes in processes and outcomes follow the introduction of this system change?	Identify behaviours or actions that might be affected, and consider these as outcomes in meta-analysis or meta-regression analyses. To account for correlations among them, multivariate methods might be considered. Consider meta-regression to examine the mediating effects of intermediate outcomes.	See section 8.
At what levels in the system are they experienced?		Identify units (eg, individuals or organisations) whose behaviour or actions might be affected, and consider these as outcomes in meta-analysis or meta-regression. Multilevel models might be appropriate to capture the different levels of impact, although may require access to individual participant data.	See section 8.

of sugar-sweetened beverages included a system-based logic model, detailing how the characteristics of the beverages, and the physiological characteristics and psychological characteristics of individuals, are thought to impact on outcomes such as weight gain and cardiovascular disease.<sup>17</sup> The logic model informs the selection of outcomes and the general plans for synthesis of the findings of included studies. However, system-based models do not usually include details of how implementation of an intervention into the system is likely to affect subsequent outcomes. They therefore have a limited role in informing syntheses that seek to explain mechanisms of action.

A quantitative synthesis may draw on a specific proposed framework for how an intervention might work; these are sometimes referred to as process-orientated logic models, and may be strongly driven by qualitative research evidence.<sup>12</sup> They represent causal processes, describing what components or aspects of an intervention are thought to impact on what behaviours and actions, and what the further consequences of these impacts are likely to be.<sup>18</sup> They may encompass mediators of effect and moderators of effect. A synthesis may simply adopt the proposed causal model at face value and attempt to quantify the relationships described therein. Where more than one possible causal model is available, a synthesis may explore which of the models is better supported by the data, for example, by examining the evidence for specific links within the model or by identifying a statistical model that corresponds to the overall causal model.<sup>18 19</sup>

A systematic review on community-level interventions for improving access to food in low-income and middle-income countries was based on a logic model that depicts how interventions might lead to improved health status.<sup>20</sup> The model includes direct effects, such as increased financial resources of individuals and decreased food prices; intermediate effects, such as increased quantity of food available and increase in intake; and main outcomes of interest, such as nutritional status and health indicators. The planned statistical synthesis, however, was to tackle these one at a time.

**Considering the types of studies available**

Studies of the effects of complex interventions may be randomised or non-randomised, and often involve clustering of participants within social or organisational units. Randomised trials, if sufficiently large, provide the most convincing evidence about the effects of interventions because randomisation should result in intervention and comparator groups with similar distributions of both observed and unobserved baseline characteristics. However, randomised trials of complex interventions may be difficult or impossible to undertake, or may be performed only in specific contexts, yielding results that are not generalisable. Non-randomised study designs include so-called ‘quasi-experiments’ and may be longitudinal studies, including interrupted time series and





**Table 2** Quantitative graphical and synthesis approaches mentioned in the paper, with their main strengths and weaknesses in the context of complex interventions

Methodological approach	Data requirements from each study	Main strengths	Main limitations
Forest plot (without overall effect)	Effect size and CI on the same metric	Widely familiar; each study clearly identified	Replication (of similar research questions) across studies is uncommon; effect size data may not be available
Albatross plot	P value, sample size and direction of effect	Data requirements are basic, so usually met; possibility of making indirect inferences on underlying effect sizes	Does not provide estimate of effect size; studies not clearly identified
Harvest plot	Conclusion of statistical test for effect; study feature(s) of interest	Data requirements are basic, so usually met; multiple outcomes can easily be displayed	Arbitrary distinction of studies according to statistical test; does not provide estimate of effect size
Effect direction plot	Conclusion of statistical test for effect; study feature(s) of interest	Data requirements are basic, so usually met; multiple outcomes can easily be displayed	Arbitrary distinction of studies according to statistical test; does not provide estimate of effect size; studies not clearly identified
Bubble plot	Conclusion of statistical analysis for effect; study feature(s) of interest	Data requirements are basic, so usually met; multiple outcomes can easily be displayed	Arbitrary distinction of studies according to result of statistical analysis; does not provide estimate of effect size; studies not clearly identified
Binomial test	Direction of effect	Data requirements are basic, so usually met	Does not provide estimate of effect size
Combining p values	P value and direction of effect	Data requirements are basic, so usually met	Does not provide estimate of effect size
Standard meta-analysis (eg, weighted average)	Effect size and CI (or equivalent) on the same metric	Widely familiar; produces effect sizes (important for decision making)	Replication (of similar research questions) across studies is uncommon; effect size data may not be available
Multiple outcomes meta-analysis (multivariate methods)	Effect size and CI (or equivalent) on the same metric for each outcome; data on correlations between outcomes	Can strengthen analysis of one outcome by 'borrowing strength' from other outcomes	Requires reasonably large number of studies for reliable results
Subgroup analysis	Effect size and CI (or equivalent) on the same metric; study feature(s) of interest	Straightforward and widely familiar; flexible approach appropriate for examining impact of context, settings, participants, intervention characteristics	Addresses one study feature at a time; requires reasonably large number of studies for reliable results; high risk of false-positive conclusions; often has low power to detect true impacts of the features examined
Meta-regression	Effect size and CI (or equivalent) on the same metric; study feature(s) of interest	Allows multiple study features to be examined together; flexible approach appropriate for examining impact of context, settings, participants, intervention characteristics and for mediating effects of intermediate outcomes	Requires reasonably large number of studies for reliable results; high risk of false-positive conclusions; often has low power to detect true impacts of the features examined
Multiple interventions meta-analysis (network meta-analysis)	Effect size and CI (or equivalent) on the same metric; category to place each intervention	Facilitates rank ordering of interventions for the outcome	Requires interventions to be grouped into (reasonably homogenous) categories; requires similar target population for all studies; requires all categories of interventions to be 'connected' in the network

Continued

**Table 2** Continued

Methodological approach	Data requirements from each study	Main strengths	Main limitations
Components-based approach to intervention complexity	Effect size and CI (or equivalent) on the same metric; components present in each intervention	Facilitates identification of most important component(s) of complex intervention	Requires reasonably large number of studies for reliable results; Assumptions required about whether components act additively or otherwise
Qualitative comparative analysis	Effect size estimates and study features of interest	Supports non-linear effects; multiple pathways to effectiveness; operates in 'small n' scenarios	Produces explanatory, rather than predictive, findings
Model-driven meta-analysis	Assumed causal model (logic model); effect size information for each relevant path in the model	Flexible approach to combining evidence; forces thinking about <i>how</i> effects arise Flexible approach to combining evidence; can supplement evidence with model-based assumptions when evidence is not available; wider focus beyond the intervention may include contextual information and dynamic interrelationships	Dependent on appropriate assumptions being made in the causal model and availability of data
Mathematical models and system science methods	Assumed model; variable data requirements		Heavily reliant on assumptions going into the model; may require very large data sets

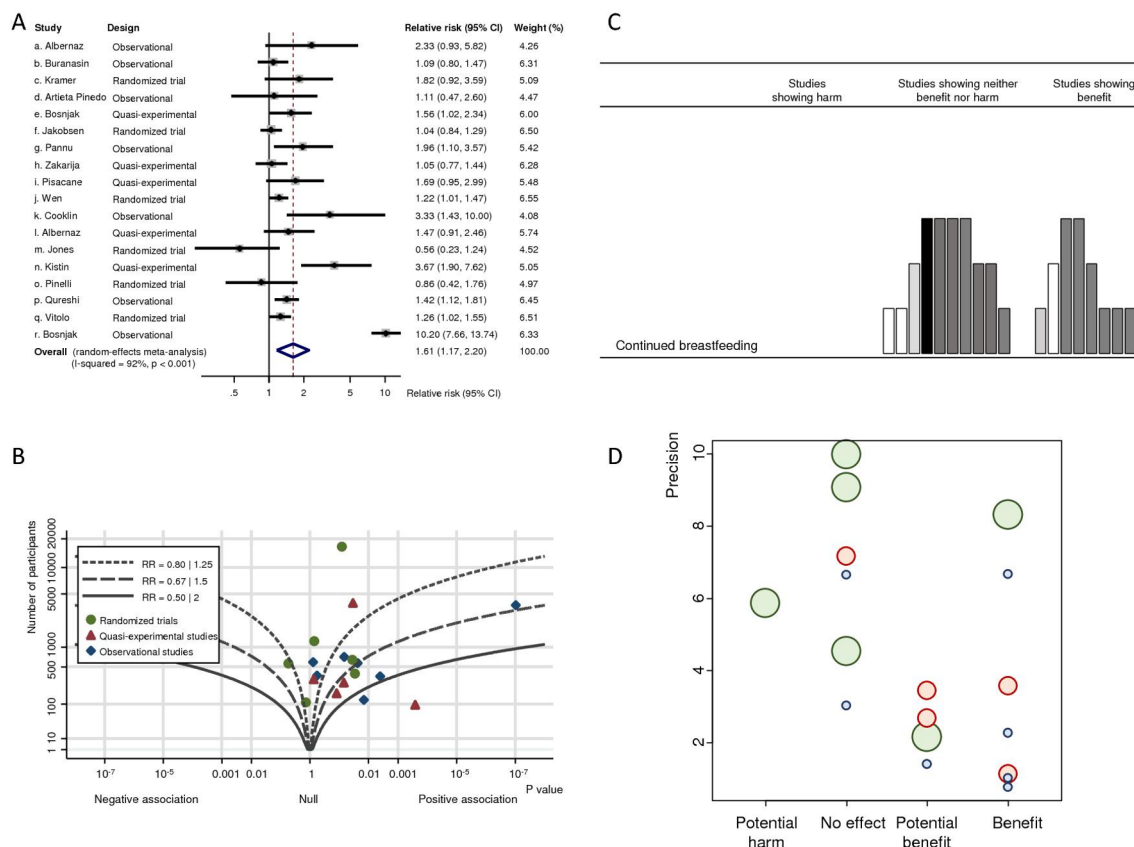
before-after studies, with or without a control group. Non-randomised studies are at greater risk of bias, sometimes substantially so, although may be undertaken in contexts that are more relevant to decision making. Analyses of non-randomised studies often use statistical controls for confounders to account for differences between intervention groups, and challenges are introduced when different sets of confounders are used in different studies.<sup>21 22</sup>

Randomised trials and non-randomised studies might both be included in a review, and analysts may have to decide whether to combine these in one synthesis, and whether to combine results from different types of non-randomised studies in a single analysis. Studies may differ in two ways: by answering different questions, or by answering similar questions with different risks of bias. The research questions must be sufficiently similar and the studies sufficiently free of bias for a synthesis to be meaningful. In the breastfeeding review, randomised, quasi-experimental and observational studies were combined; no evidence suggested that the effects differed across designs.<sup>15</sup> In practice, many methodologists generally recommend against combining randomised with non-randomised studies.<sup>23</sup>

**PREPARING FOR A QUANTITATIVE SYNTHESIS**

Before undertaking a quantitative synthesis of complex interventions, it can be helpful to begin the synthesis non-quantitatively, looking at patterns and characteristics of the data identified. Systematic tabulation of information is recommended, and this might be informed by a prespecified logic model. The most established framework for non-quantitative synthesis is that proposed by Popay *et al.*<sup>24</sup> The Cochrane Consumers and Communication group succinctly summarise the process as an 'investigation of the similarities and the differences between the findings of different studies, as well as exploration of patterns in the data'.<sup>25</sup> Another useful framework was described by Petticrew and Roberts.<sup>26</sup> They identify three stages in the initial narrative synthesis: (1) Organisation of studies into logical categories, the structure of which will depend on the purpose of the synthesis, possibly relating to study design, outcome or intervention types. (2) Within-study analysis, involving the description of findings within each study. (3) Cross-study synthesis, in which variations in study characteristics and potential biases are integrated and the range of effects described. Aspects of this process are likely to be implemented in any systematic review, even when a detailed quantitative synthesis is undertaken.

In some circumstances the available data are too diverse, too non-quantitative or too sparse for a quantitative synthesis to be meaningful even if it is possible. The best that can be achieved in many reviews of complex interventions is a non-quantitative synthesis following the guidance given in the above frameworks.



**Figure 1** Example graphical displays of data from a review of interventions to promote breast feeding, for the outcome of continued breast feeding up to 23 months.<sup>15</sup> Panel A: Forest plot for relative risk (RR) estimates from each study. Panel B: Albatross plot of p value against sample size (effect contours drawn for risk ratios assuming a baseline risk of 0.15; sample sizes and baseline risks extracted from the original papers by the current authors); Panel C: Harvest plot (heights reflect design: randomised trials (tall), quasi-experimental studies (medium), observational studies (short); bar shading reflects follow-up: longest follow-up (black) to shortest follow-up (light grey) or no information (white)). Panel D: Bubble plot (bubble sizes and colours reflect design: randomised trials (large, green), quasi-experimental studies (medium, red), observational studies (small, blue); precision defined as inverse of the SE of each effect estimate (derived from the CIs); categories are: “Potential Harm”: RR < 0.8; “No Effect”: RRs between 0.8 and 1.25; “Potential Benefit”: RR > 1.25 and CI includes RR=1; “Benefit”: RR > 1.25 and CI excludes RR=1).

## OPTIONS WHEN EFFECT SIZE ESTIMATES CANNOT BE OBTAINED OR STUDIES ARE TOO DIVERSE TO COMBINE

### Graphical approaches

Graphical displays can be very valuable to illustrate patterns in results of studies.<sup>27</sup> We illustrate some options in figure 1. Forest plots are the standard illustration of the results of multiple studies (see figure 1, panel A), but require a similar effect size estimate from each study. For studies of complex interventions, the diversity of approaches to the intervention, the context,<sup>1</sup> evaluation approaches and reporting differences can lead to considerable variation across studies in what results are available. Some novel graphical approaches have been proposed for such situations. A recent development is the albatross plot, which plots p values against sample sizes, with approximate effect-size contours superimposed (see figure 1, panel B).<sup>28</sup> The contours are computed from the p values and sample sizes, based on an assumption about the type of analysis that would have given rise to the p values. Although these plots are designed for situations when effect size estimates are not available, the

contours can be used to infer approximate effect sizes from studies that are analysed and reported in highly diverse ways. Such an advantage may prove to be a disadvantage, however, if the contours are overinterpreted.

Harvest plots have been proposed by Ogilvie *et al* as a graphical extension of a vote counting approach to synthesis (see figure 1, panel C).<sup>29</sup> However, approaches based on vote counting of statistically significant results have been criticised on the basis of their poor statistical properties, and because statistical significance is an outdated and unhelpful notion.<sup>30</sup> The harvest plot is a matrix of small illustrations, with different outcome domains defining rows and different qualitative conclusions (negative effect, no effect, positive effect) defining columns. Each study is represented by a bar that is positioned according to its measured outcome and qualitative conclusion. Bar heights and shadings can depict features of the study, such as objectivity of the outcome measure, suitability of the study design and study quality.<sup>29 31</sup> A similar idea to the harvest plot is the effect direction plot proposed by Thomson and Thomas.<sup>32</sup>

A device to plot the findings from a large and complex collection of evidence is a bubble plot (see figure 1, panel D). A bubble plot illustrates the direction of each finding (or whether the finding was unclear) on a horizontal scale, using a vertical scale to indicate the volume of evidence, and with bubble sizes to indicate some measure of credibility of each finding. Such an approach can also depict findings of collections of studies rather than individual studies, and was used successfully, for example, to summarise findings from a review of systematic reviews of the effects of acupuncture on various indications for pain.<sup>33</sup>

### Statistical methods not based on effect size estimates

We have mentioned that a frequent problem is that standard meta-analysis methods cannot be used because data are not available in a similar format from every study. In general, the core principles of meta-analysis can be applied even in this situation, as is highlighted in the *Cochrane Handbook*, by addressing the questions: 'What is the direction of effect?'; 'What is the size of effect?'; 'Is the effect consistent across studies?'; and 'What is the strength of evidence for the effect?'.<sup>34</sup>

Alternatives to the estimation of effect sizes could be used more often than they are in practice, allowing some basic statistical inferences despite diversely reported results. The most fundamental analysis is to test the overall null hypothesis of no effect in any of the studies. Such a test can be undertaken using only minimally reported information from each study. At its simplest, a binomial test can be performed using only the direction of effect observed in each study, irrespective of its CI or statistical significance.<sup>35</sup> Where exact p values are available as well as the direction of effect, a more powerful test can be performed by combining these using, for example, Fisher's combination of p values.<sup>36</sup> It is important that these p values are computed appropriately, however, accounting for clustering or matching of participants within the studies. Rejecting the null model based on such tests provides no information about the magnitude of the effect, providing information only on whether at least one study shows an effect is present, and if so, its direction.<sup>37</sup>

## STANDARD SYNTHESIS METHODS

### Meta-analysis for overall effect

Probably the most familiar approach to meta-analysis is that of estimating a single summary effect across similar studies. This simple approach lends itself to the use of forest plots to display the results of individual studies as well as syntheses, as illustrated for the breastfeeding studies in figure 1 (panel A). This analysis addresses the broad question of whether evidence from a collection of studies supports an impact of the complex intervention of interest, and requires that every study makes a comparison of a relevant intervention against a similar alternative. In the context of complex interventions, this is described

by Caldwell and Welton as the 'lumping' approach,<sup>38</sup> and by Guise *et al* as the 'holistic' approach.<sup>5 6</sup> One key limitation of the simple approach is that it requires similar types of data from each study. A second limitation is that the meta-analysis result may have limited relevance when the studies are diverse in their characteristics. Fixed-effect models, for instance, are unlikely to be appropriate for complex interventions because they ignore between-studies variability in underlying effect sizes. Results based on random-effects models will need to be interpreted by acknowledging the spread of effects across studies, for example, using prediction intervals.<sup>39</sup>

A common problem when undertaking a simple meta-analysis is that individual studies may report many effect sizes that are correlated with each other, for example, if multiple outcomes are measured, or the same outcome variable is measured at several time points. Numerous approaches are available for dealing with such multiplicity, including multivariate meta-analysis, multi-level modelling, and strategies for selecting effect sizes.<sup>40</sup> A very simple strategy that has been used in systematic reviews of complex interventions is to take the median effect size within each study, and to summarise these using the median of these effect sizes across studies.<sup>41</sup>

### Exploring heterogeneity

Diversity in the types of participants (and contexts), interventions and outcomes are key to understanding sources of complexity.<sup>9</sup> Many of these important sources of heterogeneity are most usefully examined—to the extent that they can reliably be understood—using standard approaches for understanding variability across studies, such as subgroup analyses and meta-regression.

A simple strategy to explore heterogeneity is to estimate the overall effect separately for different levels of a factor using subgroup analyses (referring to subgrouping studies rather than participants).<sup>42</sup> As an example, McFadden *et al* conducted a systematic review and meta-analysis of 73 studies of support for healthy breastfeeding mothers with healthy term babies.<sup>43</sup> They calculated separate average effects for interventions delivered by a health professional, a lay supporter or with mixed support, and found that the effect on cessation of exclusive breast feeding at up to 6 months was greater for lay support compared with professionals or mixed support ( $p=0.02$ ). Guise *et al* provide several ways of grouping studies according to their interventions, for example, grouping studies by key components, by function or by theory.<sup>5 6</sup>

Meta-regression provides a flexible generalisation to subgroup analyses, whereby study-level covariates are included in a regression model using effect size estimates as the dependent variable.<sup>44 45</sup> Both continuous and categorical covariates can be included in such models; with a single categorical covariate, the approach is essentially equivalent to subgroup analyses. Meta-regression with continuous covariates in theory allows the extrapolation of relationships to contexts that were not examined in any of the studies, but this should generally be avoided.



For example, if the effect of an interventional approach appears to increase as the size of the group to which it is applied decreases, this does not mean that it will work even better when applied to a single individual. More generally, the mathematical form of the relationship modelled in a meta-regression requires careful selection. Most often a linear relationship is assumed, but a linear relationship does not permit step changes such as might occur if an interventional approach requires a particular level of some feature of the underlying system before it has an effect.

Several texts provide guidance for using subgroup analysis and meta-regression in a general context<sup>45 46</sup> and for complex interventions.<sup>3 4 47</sup> In principle, many aspects of complexity in interventions can be addressed using these strategies, to create an understanding of the 'response surface'.<sup>48–50</sup> However, in practice, the number of studies is often too small for reliable conclusions to be drawn. In general, subgroup analysis and meta-regression are fraught with dangers associated with having few studies, many sources of variation across study features and confounding of these features with each other as well as with other, often unobserved, variables. It is therefore important to prespecify a small number of plausible sources of diversity so as to reduce the danger of reaching spurious conclusions based on study characteristics that correlate with the effects of the interventions but are not the cause of the variation. The ability of statistical analyses to identify true sources of heterogeneity will depend on the number of studies, the sizes of the studies and the true differences between effects in studies with different characteristics.

## SYNTHESIS METHODS FOR UNDERSTANDING COMPONENTS OF THE INTERVENTION

When interventions comprise distinct components, it is attractive to separate out the individual effects of these components.<sup>51</sup> Meta-regression can be used for this, using covariates to code the presence of particular features in each intervention implementation. As an example, Blakemore *et al* analysed 39 intervention comparisons from 33 independent studies aiming to reduce urgent healthcare use in adults with asthma.<sup>52</sup> Effect size estimates were coded according to components used in the interventions, and the authors found that multicomponent interventions including skills training, education and relapse prevention appeared particularly effective. In another example, of interventions to support family caregivers of people with Alzheimer's disease,<sup>53</sup> the authors used methods for decomposing complex interventions proposed by Czaja *et al*,<sup>54</sup> and created covariates that reduced the complexity of the interventions to a small number of features about the intensity of the interventions. More sophisticated models for examining components have been described by Welton *et al*,<sup>55</sup> Ivers *et al*<sup>56</sup> and Madan *et al*.<sup>57</sup>

A component-level approach may be useful when there is a need to disentangle the 'active ingredients' of an intervention, for example, when adapting an existing intervention for a new setting. However, components-based approaches require assumptions, such as whether individual components are additive or interact with each other. Furthermore, the effects of components can be difficult to estimate if they are used only in particular contexts or populations, or are strongly correlated with use of other components. An alternative approach is to treat each combination of components as a separate intervention. These separate interventions might then be compared in a single analysis using network meta-analysis. A network meta-analysis combines results from studies comparing two or more of a larger set of interventions, using indirect comparisons via common comparators to rank-order all interventions.<sup>47 58 59</sup> As an example, Achana *et al* examined the effectiveness of safety interventions on the uptake of three poisoning prevention practices in households with children. Each singular combination of intervention components was defined as a separate intervention in the network.<sup>60</sup> Network meta-analysis may also be useful when there is a need to compare multiple interventions to answer an 'in principle' question of which intervention is most effective. Consideration of the main goals of the synthesis will help those aiming to prepare guidelines to decide which of these approaches is most appropriate to their needs.

A case study exploring components is provided in [box 1](#), and an illustration is provided in [figure 2](#). The component-based analysis approach can be likened to

### Box 1 Example of understanding components of psychosocial interventions for coronary heart disease

Welton *et al* reanalysed data from a Cochrane review<sup>89</sup> of randomised controlled trials assessing the effects of psychological interventions on mortality and morbidity reduction for people with coronary heart disease.<sup>55</sup> The Cochrane review focused on the effectiveness of any psychological intervention compared with usual care, and found evidence that psychological interventions reduced non-fatal reinfarctions and depression and anxiety symptoms. The Cochrane review authors highlighted the large heterogeneity among interventions as an important limitation of their review.

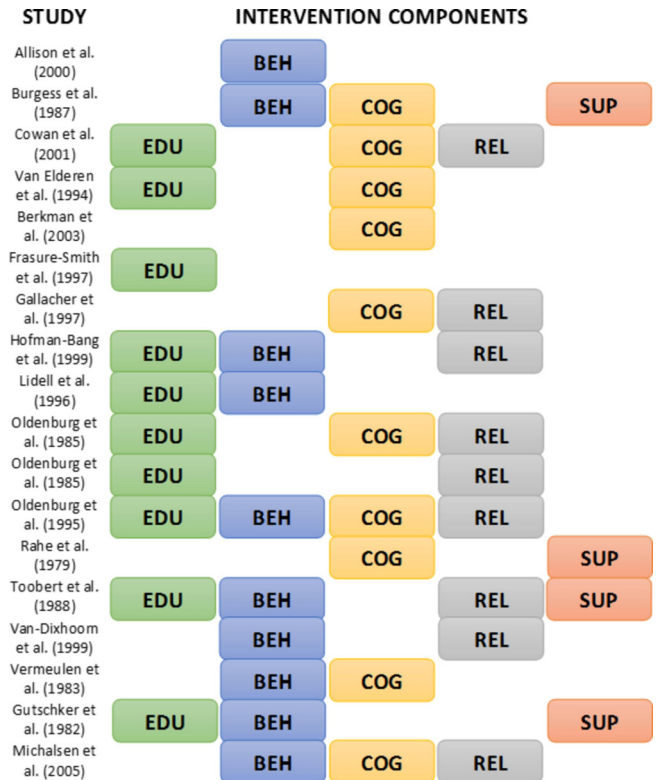
Welton *et al* were interested in the effects of the different intervention components. They classified interventions according to which of five key components were included: educational, behavioural, cognitive, relaxation and psychosocial support ([figure 2](#)). Their reanalysis examined the effect of each component in three different ways: (1) An additive model assuming no interactions between components. (2) A two-factor interaction model, allowing for interactions between pairs of components. (3) A network meta-analysis, defining each combination of components as a separate intervention, therefore allowing for full interaction between components. Results suggested that interventions with behavioural components were effective in reducing the odds of all-cause mortality and non-fatal myocardial infarction, and that interventions with behavioural and/or cognitive components were effective for reducing depressive symptoms.

a factorial trial, in that it attempts to separate out the effects of individual components of the complex interventions, and the network meta-analysis approach can be likened to a multiarm trial approach, where each complex intervention in the set of studies is a different arm in the trial.<sup>47</sup> Deciding between the two approaches can leave the analyst caught between the need to ‘split’ components to reflect complexity (and minimise heterogeneity) and ‘lump’ to make an analysis feasible. Both approaches can be used to examine other features of interventions, including interventions designed for delivery at different levels. For example, a review of the effects of interventions for children exposed to domestic violence and abuse included studies of interventions targeted at children alone, parents alone, children and parents together, and parents and children separately.<sup>61</sup> A network meta-analysis approach was taken to the synthesis, with the people targeted by the intervention used as a distinguishing feature of the interventions included in the network.

A common limitation when implementing these quantitative methods in the context of complex interventions is that replication of the same intervention in two or more studies is rare. Qualitative comparative analysis (QCA) might overcome this problem, being designed to address the ‘small N; many variables’ problem.<sup>62</sup> QCA involves: (1) Identifying theoretically driven thresholds for determining intervention success or failure. (2) Creating a ‘truth table’, which takes the form of a matrix, cross-tabulating all possible combinations of conditions (eg, participant and intervention characteristics) against each study and its associated outcomes. (3) Using Boolean algebra to eliminate redundant conditions and to identify configurations of conditions that are necessary and/or sufficient to trigger intervention success or failure. QCA can usefully complement quantitative integration, sometimes in the context of synthesising diverse types of evidence.

**SYNTHESIS METHODS FOR UNDERSTANDING MECHANISMS OF ACTION**

An alternative purpose of a synthesis is to gain insight into the mechanisms of action behind an intervention, to inform its generalisability or applicability to a particular context. Such syntheses of quantitative data may complement syntheses of qualitative data,<sup>11</sup> and the two forms might be integrated.<sup>12</sup> Logic models, or theories of action, are important to motivate investigations of mechanism. The synthesis is likely to focus on intermediate outcomes reflecting intervention processes, and on mediators of effect (factors that influence how the intervention affects an outcome measure). Two possibilities for analysis are to use these intermediate measurements as predictors of main outcomes using meta-regression methods,<sup>63</sup> or to use multivariate meta-analysis to model the intermediate and main outcomes simultaneously, exploiting and estimating the correlations between them.<sup>64 65</sup> If the synthesis suggests that hypothesised chains of outcomes



**Figure 2** Intervention components in the studies integrated by Welton *et al* (a sample of 18 from 56 active treatment arms). EDU, educational component; BEH, behavioural component; COG, cognitive component; REL, relaxation component; SUP, psychosocial support component.

hold, this lends weight to the theoretical model underlying the hypothesis.

An approach to synthesis closely identified with this category of interventions is model-driven meta-analysis, in which different sources of evidence are integrated within a causal path model akin to a directed acyclic graph. A model-driven meta-analysis is an explanatory analysis.<sup>66</sup> It attempts to go further than a standard meta-analysis or meta-regression to explore how and why an intervention works, for whom it works, and which aspects (of the intervention (factors) are driving overall effect. Such syntheses have been described in frequentist<sup>19 67-70</sup> and Bayesian<sup>71 72</sup> frameworks and are variously known as model-driven meta-analysis, linked meta-analysis, meta-mediation analysis and meta-analysis of structural equation models. In their simplest form, standard meta-analyses estimate a summary correlation independently for each pair of variables in the model. The approach is inherently multivariate, requiring the estimation of multiple correlations (which, if obtained from a single study, are also not independent).<sup>73-75</sup> Each study is likely to contribute fragments of the correlation matrix. A summary correlation matrix, combined either by fixed-effects or random-effects methods, then serves as the input for subsequent analysis via a standardised regression or structural equation model.

## Box 2 Example of a model-driven meta-analysis for type 2 diabetes

Brown *et al* present a model-driven meta-analysis of correlational research on psychological and motivational predictors of diabetes outcomes, with medication and dietary adherence factors as mediators.<sup>76</sup> In a linked methodological paper, they present the a priori theoretical model on which their analysis is based.<sup>68</sup> The model is simplified in figure 3, and summarised for the dietary adherence pathway only. The aim of their full analysis was to determine the predictive relationships among psychological factors and motivational factors on metabolic control and body mass index (BMI), and the role of behavioural factors as possible mediators of the associations among the psychological and motivational factors and metabolic control and BMI outcomes.

The analysis is based on a comprehensive systematic review. Due to the number of variables in their full model, 775 individual correlational or predictive studies reported across 739 research papers met eligibility criteria. Correlations between each pair of variables in the model were summarised using an overall average correlation, and homogeneity assessed. Multivariate analyses were used to estimate a combined correlation matrix. These results were used, in turn, to estimate path coefficients for the predictive model and their standard errors. For the simplified model illustrated here, the results suggested that coping and self-efficacy were strongly related to dietary adherence, which was strongly related to improved glycaemic control and, in turn, a reduction in diabetic complications.

An example is provided in box 2. The model in figure 3 postulates that the effect of ‘Dietary adherence’ on ‘Diabetes complications’ is not direct but is mediated by ‘Metabolic control’.<sup>76</sup> The potential for model-driven meta-analysis to incorporate such indirect effects also allows for mediating effects to be explicitly tested and in so doing allows the meta-analyst to identify and explore the mechanisms underpinning a complex intervention.<sup>77</sup>

### SYNTHESIS APPROACHES FOR UNDERSTANDING COMPLEXITIES OF THE SYSTEM

Syntheses may seek to address complexities of the system to understand either the impact of the system on the effects of the intervention or the effects of the intervention on the system. This may start by modelling the salient features of the system’s dynamics, rather than focusing on interventions. Subgroup analysis and meta-regression are useful approaches for investigating the extent to which an intervention’s effects depend on baseline features of the system, including aspects of the context. Sophisticated meta-regression models might investigate multiple

baseline features, using similar approaches to the component-based meta-analyses described earlier. Specifically, aspects of context or population characteristics can be regarded as ‘components’ of the system into which the intervention is introduced, and similar statistical modelling strategies used to isolate effects of individual factors, or interactions between them.

When interventions act at multiple levels, it may be important to understand the effects at these different levels. Outcomes may be measured at different levels (eg, at patient, clinician and clinical practice levels) and analysed separately. Qualitative research plays a particularly important role in identifying the outcomes that should be assessed through quantitative synthesis.<sup>12</sup> Care is needed to ensure that the unit of analysis issues are addressed. For example, if clinics are the unit of randomisation, then outcomes measured at the clinic level can be analysed using standard methods, whereas outcomes measured at the level of the patient within the clinic would need to account for clustering. In fact, multiple dependencies may arise in such data, when patients receive care in small groups. Detailed investigations of effect at different levels, including interactions between the levels, would lend themselves to multilevel (hierarchical) models for synthesis. Unfortunately, individual participant data at all levels of the hierarchy are needed for such analyses.

Model-based approaches also offer possibilities for addressing complex systems; these include economic models, mathematical models and systems science methods generally.<sup>78–80</sup> Broadly speaking, these provide mathematical representations of logic models, and analyses may involve incorporation of empirical data (eg, from systematic reviews), computer simulation, direct computation or a mixture of these. Multiparameter evidence synthesis methods might be used.<sup>81 82</sup> Approaches include models to represent systems (eg, systems dynamics models) and approaches that simulate individuals within the system (eg, agent-based models).<sup>79</sup> Models can be particularly useful when empirical evidence does not address all important considerations, such as ‘real-world’ contexts, long-term effects, non-linear effects and complexities such as feedback loops and threshold effects. An example of a model-based approach to synthesis is provided in box 3. The challenge when adopting these approaches is often in the identification of system components, and accurately estimating causes and effects (and uncertainties). There are few examples of the use of these analytical tools in systematic

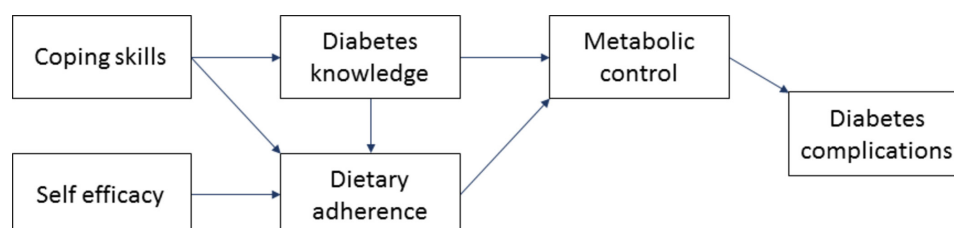


Figure 3 Theoretical diabetes care model (adapted from Brown *et al*<sup>68</sup>).



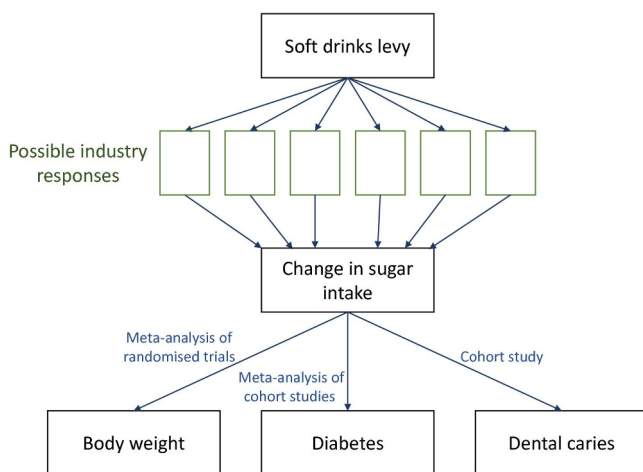
**Box 3 Example of a mathematical modelling approach for soft drinks industry levy**

Briggs *et al* examined the potential impact of a soft drinks levy in the UK, considering possible different types of response to the levy by industry.<sup>90</sup> Various scenarios were posited, with effects on health outcomes informed by empirical data from randomised trials and cohort studies of association between sugar intake and body weight, diabetes and dental caries. Figure 4 provides a simple characterisation of how the empirical data were fed into the model. Inputs into the model included levels of consumption of various types of drinks (by age and sex), volume of drinks sales, and baseline levels of obesity, diabetes and dental caries (by age and sex). The authors concluded that health gains would be greatest if industry reacted by reformulating their products to include less sugar.

reviews, but they may be useful when the focus of analysis is on understanding the causes of complexity in a given system rather than on the impact of an intervention.

**CONSIDERATIONS OF BIAS AND RELEVANCE**

It is always important to consider the extent to which (1) The findings from each study have internal validity, particularly for non-randomised studies which are typically at higher risk of bias. (2) Studies may have been conducted but not reported because of unexciting findings. (3) Each study is applicable to the purposes of the review, that is, has external validity (or ‘directness’), in the language of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group.<sup>83</sup> At minimum, internal and external validity should be examined and reported, and the risk of publication bias assessed, and these can be achieved through the GRADE framework.<sup>10</sup> With sufficient studies, information collected might be used in meta-regression analyses to evaluate empirically whether studies with and without specific sources of bias or indirectness differ in their results.



**Figure 4** Simplified version of the conceptual model used by Briggs *et al* (adapted from Briggs *et al*<sup>90</sup>).

It may be desirable to learn about a specific setting, intervention type or outcome measure more directly than others. For example, to inform a decision for a low-income setting, emphasis should be placed on results of studies performed in low-income countries. One option is to restrict the synthesis to these studies. An alternative is to model the dependence of an intervention’s effect on some feature(s) related to the income setting, and extract predictions from the model that are most relevant to the setting of interest. This latter approach makes fuller use of available data, but relies on stronger assumptions.

Often, however, the accumulated studies are too few or too disparate to draw conclusions about the impact of bias or relevance. On rare occasions, syntheses might implement formal adjustments of individual study results for likely biases. Such adjustments may be made by imposing prior distributions to depict the magnitude and direction of any biases believed to exist.<sup>84 85</sup> The choice of a prior distribution may be informed by formal assessments of risk of bias, by expert judgement, or possibly by empirical data from meta-epidemiological studies of biases in randomised and/or non-randomised studies.<sup>86</sup> For example, Wolf *et al* implemented a prior distribution based on findings of a meta-epidemiological study<sup>87</sup> to adjust for lack of blinding in studies of interventions to improve quality of point-of-use water sources in low-income and middle-income settings.<sup>88</sup> Unfortunately, empirical evidence of bias is mostly limited to clinical trials, is weak for trials of public health and social care interventions, and is largely non-existent for non-randomised studies.

**CONCLUSION**

Our review of quantitative synthesis methods for evaluating the effects of complex interventions has outlined many possible approaches that might be considered by those collating evidence in support of guideline development. We have described three broad categories: (1) Non-quantitative methods, including tabulation, narrative and graphical approaches. (2) Standard meta-analysis methods, including meta-regression to investigate study-level moderators of effect. (3) More advanced synthesis methods, in which models allow exploration of intervention components, investigation of both moderators and mediators, examination of mechanisms, and exploration of complexities of the system.

The choice among these approaches will depend on the purpose of the synthesis, the similarity of the studies included in the review, the level of detail available from the studies, the nature of the results reported in the studies, the expertise of the synthesis team, and the resources available. Clearly the advanced methods require more expertise and resources than the simpler methods. Furthermore, they require a greater level of detail and typically a sizeable evidence base. We therefore expect them to be used seldomly; our aim here is largely to articulate what they can achieve so that they



can be adopted when they are appropriate. Notably, the choice among these approaches will also depend on the extent to which guideline developers and users at global, national or local levels understand and are willing to base their decisions on different methods. Where possible, it will thus be important to involve concerned stakeholders during the early stages of the systematic review process to ensure the relevance of its findings.

Complexity is common in the evaluation of public health interventions at individual, organisational or community levels. To help systematic review and guideline development teams decide how to address this complexity in syntheses of quantitative evidence, we summarise considerations and methods in tables 1 and 2. We close with the important remark that quantitative synthesis is not always a desirable feature of a systematic review. Whereas some sophisticated methods are available to deal with a variety of complex problems, on many occasions—perhaps even the majority in practice—the studies may be too different from each other, too weak in design or have data too sparse, for statistical methods to provide insight beyond a commentary on what evidence has been identified.

#### Author affiliations

<sup>1</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

<sup>2</sup>Department of Educational Psychology and Learning Systems, College of Education, Florida State University, Tallahassee, Florida, USA

<sup>3</sup>Clinical Epidemiology Program, Ottawa Hospital Research Institute, The Ottawa Hospital, Ottawa, Ontario, Canada

<sup>4</sup>Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada

<sup>5</sup>NIHR Collaboration for Leadership in Applied Health Care (CLAHRC) West, University Hospitals Bristol NHS Foundation Trust, Bristol, UK

<sup>6</sup>Institute for Medical Information Processing, Biometry and Epidemiology, Pettenkofer School of Public Health, LMU Munich, Munich, Germany

<sup>7</sup>EPPI-Centre, Department of Social Science, University College London, London, UK

**Acknowledgements** The authors thank the following for helpful comments on earlier drafts of the paper: Philippa Easterbrook, Matthias Egger, Anayda Portela, Susan L Norris, Mark Petticrew.

**Contributors** JPTH co-led the project, conceived the paper, led discussions and wrote the first draft. JAL-L undertook analyses, contributed to discussions and contributed to writing the manuscript. BJB drafted material on mechanisms, contributed to discussions and contributed extensively to writing the manuscript. SRD screened and categorised the results of the literature searches, collated examples and contributed to discussions. SD undertook searches to identify relevant literature and contributed to discussions. JMG contributed to discussions and commented critically on drafts. LAM undertook analyses, contributed to discussions and commented critically on drafts. THMM contributed examples, contributed to discussions and commented critically on drafts. EAR and JT contributed to discussions and commented critically on drafts. DMC co-led the project, contributed to discussions and drafted extensive parts of the paper. All authors approved the final version of the manuscript.

**Funding** Funding provided by the World Health Organization Department of Maternal, Newborn, Child and Adolescent Health through grants received from the United States Agency for International Development and the Norwegian Agency for Development Cooperation. JPTH was funded in part by Medical Research Council (MRC) grant MR/M025209/1, by the MRC Integrative Epidemiology Unit at the University of Bristol (MC\_UU\_12013/9) and by the MRC ConDuCT-II Hub (Collaboration and innovation for Difficult and Complex randomised controlled Trials In Invasive procedures – MR/K025643/1). BJB was funded in part by grant DRL-1252338 from the US National Science Foundation (NSF). JMG holds a Canada Research Chair in Health Knowledge Transfer and Uptake. LAM is funded by a National Institute for Health Research (NIHR) Systematic Review Fellowship (RM-SR-2016-07 26). THMM was funded by the NIHR Collaboration for Leadership

in Applied Health Research and Care West (NIHR CLAHRC West). JT is supported by the NIHR Collaboration for Leadership in Applied Health Research and Care North Thames at Bart's Health NHS Trust. DMC was funded in part by NIHR grant PHR 15/49/08 and by the Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHER –MR/K0232331/1).

**Disclaimer** The views expressed are those of the authors and not necessarily those of the CRC program, the MRC, the NSF, the NHS, the NIHR or the UK Department of Health.

**Competing interests** JMG reports personal fees from the Campbell Collaboration. EAR reports being a Methods Editor with Cochrane Public Health.

**Patient consent** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data statement** No additional data are available.

**Open access** This is an open access article distributed under the terms of the Creative Commons Attribution-Non commercial IGO License (CC BY-NC 3.0 IGO), which permits use, distribution, and reproduction for non-commercial purposes in any medium, provided the original work is properly cited. In any reproduction of this article there should not be any suggestion that WHO or this article endorse any specific organization or products. The use of the WHO logo is not permitted. This notice should be preserved along with the article's original URL.

## REFERENCES

- Petticrew M, Knai C, Thomas J, *et al.* Implications of a complexity perspective for systematic reviews and guideline development in health decision making. *BMJ Glob Health* 2019;0:e000899. doi:10.1136/bmjgh-2018-000899.
- Petticrew M, Anderson L, Elder R, *et al.* Complex interventions and their implications for systematic reviews: a pragmatic approach. *J Clin Epidemiol* 2013;66:1209–14.
- Petticrew M, Rehfues E, Noyes J, *et al.* Synthesizing evidence on complex interventions: how meta-analytical, qualitative, and mixed-method approaches can contribute. *J Clin Epidemiol* 2013;66:1230–43.
- Pigott T, Shepperd S. Identifying, documenting, and examining heterogeneity in systematic reviews of complex interventions. *J Clin Epidemiol* 2013;66:1244–50.
- Guise JM, Chang C, Viswanathan M, *et al.* *Systematic reviews of complex multicomponent health care interventions (AHRQ Publication No. 14-EHC003-EF)*. Rockville, MD: Agency for Healthcare Research and Quality, 2014.
- Guise JM, Chang C, Viswanathan M, *et al.* Agency for Healthcare Research and Quality Evidence-based Practice Center methods for systematically reviewing complex multicomponent health care interventions. *J Clin Epidemiol* 2014;67:1181–91.
- World Health Organization. *WHO Handbook for Guideline Development*. 2nd edn. Geneva, Switzerland: World Health Organization, 2014.
- Rehfues EA, Stratil JM, Scheel IB, *et al.* The WHO-INTEGRATE evidence to decision framework version 1.0: integrating WHO norms and values and a complexity perspective. *BMJ Glob Health* 2019;0:e000844. doi:10.1136/bmjgh-2018-000844.
- Booth A, Moore G, Flemming K, *et al.* Taking account of context in systematic reviews and guidelines considering a complexity perspective. *BMJ Glob Health* 2019;0:e000840. doi:10.1136/bmjgh-2018-000840.
- Montgomery P, Movsisyan A, Grant SP, *et al.* Considerations of complexity in rating certainty of evidence in systematic reviews: a primer on using the grade approach in global health. *BMJ Glob Health* 2019;0:e000848. doi:10.1136/bmjgh-2018-000848.
- Flemming K, Booth A, Garside R, *et al.* Qualitative evidence synthesis for complex interventions and guideline development: clarification of the purpose, designs and relevant methods. *BMJ Glob Health* 2019;0:e000882. doi:10.1136/bmjgh-2018-000882.
- Noyes J, Booth A, Moore G, *et al.* Synthesizing quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. *BMJ Glob Health* 2019;0:e000893. doi:10.1136/bmjgh-2018-000893.
- World Health Organization. *Guideline: protecting, promoting and supporting breastfeeding in facilities providing maternity and newborn services*. Geneva: World Health Organization, 2017.
- Richardson WS, Wilson MC, Nishikawa J, *et al.* The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995;123:A12–13.

15. Sinha B, Chowdhury R, Sankar MJ, *et al*. Interventions to improve breastfeeding outcomes: a systematic review and meta-analysis. *Acta Paediatr* 2015;104:114–34.
16. Collins D, Johnson K, Becker BJ. A meta-analysis of direct and mediating effects of community coalitions that implemented science-based substance abuse prevention interventions. *Subst Use Misuse* 2007;42:985–1007.
17. von Philipsborn P, Stratil JM, Burns J, *et al*. Environmental interventions to reduce the consumption of sugar-sweetened beverages and their effects on health. *Cochrane Database Syst Rev* 2016;135:CD012292.
18. Joffe M, Mindell J. Complex causal process diagrams for analyzing the health impacts of policy interventions. *Am J Public Health* 2006;96:473–9.
19. Becker BJ. Model-based meta-analysis. In: Cooper H, Hedges LV, Valentine JC, *The Handbook of Research Synthesis and Meta-analysis*. 2nd edn. New York: Russell Sage Foundation, 2009:377–95.
20. Durao S, Schoonees A, Romokolo V. Community-level interventions for improving access to food in low- and middle-income countries (Protocol). *Cochrane Database Syst Rev* 2015;2:CD011504.
21. Aloe AM, Becker BJ, Duvendack M, *et al*. Quasi-experimental study designs series-paper 9: collecting data from quasi-experimental studies. *J Clin Epidemiol* 2017;89:77–83.
22. Becker BJ, Aloe AM, Duvendack M, *et al*. Quasi-experimental study designs series-paper 10: synthesizing evidence for effects collected from quasi-experimental studies presents surmountable challenges. *J Clin Epidemiol* 2017;89:84–91.
23. Reeves BC, Deeks JJ, Higgins JPT, *et al*. Including non-randomized studies. In: Higgins JPT, Green S, eds. *Cochrane Handbook for systematic reviews of interventions*. Chichester (UK): John Wiley & Sons, 2008.
24. Popay J, Roberts H, Sowden A. *Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC Methods Programme*. Lancaster, UK: Lancaster University, 2006.
25. Ryan R. Cochrane Consumers and Communication Review Group. 2016. Cochrane consumers and communication review group: data synthesis and analysis. <http://cccrg.cochrane.org> (accessed 1 Nov 2017).
26. Petticrew M, Roberts H. *Systematic reviews in the social sciences: a practical guide*. Malden (MA): Blackwell, 2006.
27. Anzures-Cabrera J, Higgins JPT. Graphical displays for meta-analysis: an overview with suggestions for practice. *Res Synth Methods* 2010;1:66–80.
28. Harrison S, Jones HE, Martin RM, *et al*. The albatross plot: a novel graphical tool for presenting results of diversely reported studies in a systematic review. *Res Synth Methods* 2017;8:281–9.
29. Ogilvie D, Fayer D, Petticrew M, *et al*. The harvest plot: a method for synthesising evidence about the differential effects of interventions. *BMC Med Res Methodol* 2008;8:8:8.
30. Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? *BMJ* 2001;322:226–31.
31. Crowther M, Avenell A, MacLennan G, *et al*. A further use for the Harvest plot: a novel method for the presentation of data synthesis. *Res Synth Methods* 2011;2:79–83.
32. Thomson HJ, Thomas S. The effect direction plot: visual display of non-standardised effects across multiple outcome domains. *Res Synth Methods* 2013;4:95–101.
33. Hempel S, Taylor SL, Solloway M. *Evidence map of acupuncture. VAESP Project #05-226*, 2013.
34. Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. Chichester (UK): John Wiley & Sons, 2008: 243–96.
35. Bushman BJ, Wang MC. Vote-counting procedures in meta-analysis. In: Cooper H, Hedges LV, Valentine JC, eds. *The handbook of research synthesis and meta-analysis*. 2nd edn. New York: Russell Sage Foundation, 2009: 207–20.
36. Becker BJ. Combining significance levels. In: Cooper H, Hedges LV, eds. *The handbook of research synthesis*. New York: Russell Sage Foundation, 1994: 215–30.
37. Becker BJ. Applying tests of combined significance in meta-analysis. *Psychol Bull* 1987;102:164–71.
38. Caldwell DM, Welton NJ. Approaches for synthesising complex mental health interventions in meta-analysis. *Evid Based Ment Health* 2016;19:16–21.
39. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009;172:137–59.
40. López-López JA, Page MJ, Lipsey MW, *et al*. Dealing with effect size multiplicity in systematic reviews and meta-analyses. *Res Synth Methods* 2018;68.
41. Grimshaw JM, Thomas RE, MacLennan G, *et al*. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004;8:1–72.
42. Borenstein M, Higgins JP. Meta-analysis and subgroups. *Prev Sci* 2013;14:134–43.
43. McFadden A, Gavine A, Renfrew MJ, *et al*. Support for healthy breastfeeding mothers with healthy term babies. *Cochrane Database Syst Rev* 2017;2:CD001141.
44. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002;21:589–624.
45. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21:1559–73.
46. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992;116:78–84.
47. Melendez-Torres GJ, Bonell C, Thomas J. Emergent approaches to the meta-analysis of multiple heterogeneous complex interventions. *BMC Med Res Methodol* 2015;15:47.
48. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet* 1998;351:123–7.
49. Lenz M, Steckelberg A, Richter B, *et al*. Meta-analysis does not allow appraisal of complex interventions in diabetes and hypertension self-management: a methodological review. *Diabetologia* 2007;50:1375–83.
50. Rubin DB. Meta-analysis: literature synthesis or effect-size surface estimation? *J Educ Stat* 1992;17:363–74.
51. Pesseau J, Ivers NM, Newham JJ, *et al*. Using a behaviour change techniques taxonomy to identify active ingredients within trials of implementation interventions for diabetes care. *Implement Sci* 2015;10:55.
52. Blakemore A, Dickens C, Anderson R, *et al*. Complex interventions reduce use of urgent healthcare in adults with asthma: systematic review with meta-regression. *Respir Med* 2015;109:147–56.
53. Belle SH, Czaja SJ, Schulz R, *et al*. Using a new taxonomy to combine the uncombinable: integrating results across diverse interventions. *Psychol Aging* 2003;18:396–405.
54. Czaja SJ, Schulz R, Lee CC, *et al*. A methodology for describing and decomposing complex psychosocial and behavioral interventions. *Psychol Aging* 2003;18:385–95.
55. Welton NJ, Caldwell DM, Adamopoulos E, *et al*. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. *Am J Epidemiol* 2009;169:1158–65.
56. Ivers N, Tricco AC, Trikalinos TA, *et al*. Seeing the forests and the trees-innovative approaches to exploring heterogeneity in systematic reviews of complex interventions to enhance health system decision-making: a protocol. *Syst Rev* 2014;3:88.
57. Madan J, Chen Y-F, Aveyard P, *et al*. Synthesis of evidence on heterogeneous interventions with multiple outcomes recorded over multiple follow-up times reported inconsistently: a smoking cessation case-study. *J R Stat Soc Ser A Stat Soc* 2014;177:295–314.
58. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;331:897–900.
59. Grant ES, Calderbank-Batista T. Network meta-analysis for complex social interventions: problems and potential. *J Soc Social Work Res* 2013;4:406–20.
60. Achana FA, Cooper NJ, Bujkiewicz S, *et al*. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Med Res Methodol* 2014;14:92.
61. Howarth E, Moore THM, Welton NJ, *et al*. IMPROving Outcomes for children exposed to domestic Violence (IMPROVE): an evidence synthesis. *Public Health Res* 2016;4:1–342.
62. Thomas J, O'Mara-Eves A, Brunton G. Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example. *Syst Rev* 2014;3:67.
63. Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Stat Methods Med Res* 1993;2:173–92.
64. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Stat Med* 2011;30:2481–98.
65. Raudenbush SW, Becker BJ, Kalaian H. Modeling multivariate effect sizes. *Psychol Bull* 1988;103:111–20.
66. Cook TD, Cooper H, Cordray DS. *Meta-analysis for explanation: a casebook*. New York: Russell Sage Foundation, 1994.
67. Becker BJ. Meta-analysis and models of substance abuse prevention. *NIDA Res Monogr* 1997;170:96–119.

68. Brown SA, Becker BJ, García AA, *et al*. Model-driven meta-analyses for informing health care: a diabetes meta-analysis as an exemplar. *West J Nurs Res* 2015;37:517–35.
69. Lipsey MW. Using linked meta-analysis to build policy models. *NIDA Res Monogr* 1997;170:216–33.
70. Cheung MW, Chan W. Meta-analytic structural equation modeling: a two-stage approach. *Psychol Methods* 2005;10:40–64.
71. Watson SI, Lilford RJ. Integrating multiple sources of evidence: a Bayesian perspective. In: Raine R, Fitzpatrick R, Barratt H, Bevan G, Black N, Boaden R, eds. *Challenges, solutions and future directions in the evaluation of service innovations in health care and public health*. Health Services Delivery Research, 2016:1–18.
72. Stewart GB, Mengersen K, Meader N. Potential uses of Bayesian networks as tools for synthesis of systematic reviews of complex interventions. *Res Synth Methods* 2014;5:1–12.
73. Cheung MW, Hafdahl AR. Special issue on meta-analytic structural equation modeling: introduction from the guest editors. *Res Synth Methods* 2016;7:112–20.
74. Becker BJ. Using results from replicated studies to estimate linear models. *Journal of Educational Statistics* 1992;17:341–62.
75. Becker BJ. Corrections to "Using results from replicated studies to estimate linear models". *J Stat Educ* 1995;20:100–2.
76. Brown SA, García AA, Brown A, *et al*. Biobehavioral determinants of glycemic control in type 2 diabetes: a systematic review and meta-analysis. *Patient Educ Couns* 2016;99:1558–67.
77. Dunn G, Emsley R, Liu H, *et al*. Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: a methodological research programme. *Health Technol Assess* 2015;19:1–116.
78. Egger M, Johnson L, Althaus C, *et al*. Developing WHO guidelines: time to formally include evidence from mathematical modelling studies. *F1000Res* 2017;6:1584.
79. Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. *Annu Rev Public Health* 2012;33:357–76.
80. Greenwood-Lee J, Hawe P, Nettel-Aguirre A, *et al*. Complex intervention modelling should capture the dynamics of adaptation. *BMC Med Res Methodol* 2016;16:51.
81. Ades AE, Welton NJ, Caldwell D, *et al*. Multiparameter evidence synthesis in epidemiology and medical decision-making. *J Health Serv Res Policy* 2008;13(Suppl 3):12–22.
82. Colbourn T, Asseburg C, Bojke L, *et al*. Prenatal screening and treatment strategies to prevent group B streptococcal and other bacterial infections in early infancy: cost-effectiveness and expected value of information analyses. *Health Technol Assess* 2007;11:1–226.
83. Guyatt GH, Oxman AD, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
84. Turner RM, Spiegelhalter DJ, Smith GCS, *et al*. Bias modelling in evidence synthesis. *J R Stat Soc Ser A Stat Soc* 2009;172:21–47.
85. Welton NJ, Ades AE, Carlin JB, *et al*. Models for potentially biased evidence in meta-analysis using empirically based priors. *J R Stat Soc Ser A Stat Soc* 2009;172:119–36.
86. Sterne JA, Jüni P, Schulz KF, *et al*. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;21:1513–24.
87. Savović J, Jones HE, Altman DG, *et al*. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429–38.
88. Wolf J, Prüss-Ustün A, Cumming O, *et al*. Assessing the impact of drinking water and sanitation on diarrhoeal disease in low- and middle-income settings: systematic review and meta-regression. *Trop Med Int Health* 2014;19:928–42.
89. Rees K, Bennett P, West R, *et al*. Psychological interventions for coronary heart disease. *Cochrane Database Syst Rev* 2004;2:CD002902.
90. Briggs ADM, Mytton OT, Kehlbacher A, *et al*. Health impact assessment of the UK soft drinks industry levy: a comparative risk assessment modelling study. *Lancet Public Health* 2017;2:e15–e22.