Approaches for synthesising complex mental health interventions in meta-analysis
Deborah M Caldwell, Nicky J Welton

School of Social and Community Medicine, University of Bristol, Bristol, UK
Correspondence to Dr Deborah M Caldwell, d.m.caldwell@bristol.ac.uk

ABSTRACT
Clinical and statistical heterogeneity are commonplace in meta-analysis of mental health interventions. One possible source of this heterogeneity is the complexity of the intervention being evaluated. Complexity may relate to the intervention, or to the way in which it is implemented; however, the most common interpretation of a complex intervention is one which has multiple, potentially interacting components. In this article we outline different analytical strategies suggested for incorporating intervention complexity in a meta-analysis.

INTRODUCTION
According to the Centre for Evidence-Based Medicine’s hierarchies of evidence, systematic reviews of homogenous well-conducted randomised controlled trials (RCTs) provide the best evidence for evaluating intervention effectiveness. Homogeneity implies that each study included in a meta-analysis is estimating a single, true underlying relative treatment effect, so that any differences in estimates between studies are due to sampling error alone. If homogeneity does not hold, a single, fixed treatment effect meta-analysis model should not be assumed and a random effects model may be more appropriate. In systematic reviews of interventions for mental health, it could be argued that the homogeneity assumption is unlikely to hold in general. This is because the clinical variation observed across patient populations, therapist fidelity, intervention and comparator conditions, and outcomes (both patient and clinician reported) can give rise to statistical heterogeneity in meta-analysis. Indeed, heterogeneity might be considered inevitable.

In systematic reviews of mental health interventions, the presence of statistical heterogeneity may be attributable to the complexity of the intervention being evaluated leading to potentially important differences across studies. ‘Complexity’ itself is a contested term, however the Medical Research Council (MRC) have described the characteristics of complex interventions as having:

- A number of interacting components within the experimental and control interventions,
- A number and difficulty of behaviours required by those delivering or receiving the intervention,
- A number of groups or organisational levels targeted by the intervention,
- A number and variability of outcomes,
- A degree of flexibility or tailoring of the intervention permitted.

These characteristics, or any not, may be present in every complex intervention. Petticrew et al. classify characteristics of complexity as (1) those which relate to the intervention itself (such as multiple interacting components and flexibility of implementation) and (2) those which relate to the interventions’ causal pathway (such as interaction with context, multiple mediators and moderators of effect).

Strategies to handle complex interventions in meta-analysis range from ‘lumping’ all interventions together to sophisticated statistical modelling techniques. The aim of this paper is to give an overview of the different analytical strategies suggested for incorporating intervention complexity in a systematic review, illustrated using a subset of studies from a Cochrane review examining psychological therapies for reducing depressive symptoms postcoronary heart disease.

We consider complexity only as it relates to the intervention and conceptualise a complex intervention as one, which has multiple, potentially interacting components. This is the most common interpretation. Interested readers are referred to a special edition of the Journal of Clinical Epidemiology for consideration of strategies for handling to other aspects of complexity in systematic reviews.

FORMULATING THE RESEARCH QUESTION—LUMPING OR SPLITTING?
The analytical strategy for synthesising complex interventions should be pre-specified and begins with the formulation of a sensible research question, which in turn depends on the purpose of the review. The specification of a review’s objectives shapes whether the analytical strategy will ‘lump’ or ‘split’ interventions. For example, in principle research questions such as ‘Do psychological therapies (as a whole), reduce depression after coronary heart disease?’ might take a lumping approach to analysis, since this question seeks to understand effectiveness in general. However, when complex interventions are ‘lumped’ together to form a single comparator, any between interventions variation is masked and is likely to manifest as increased, but unexplained, heterogeneity. Of course, the decision to lump interventions may also be taken for practical reasons, such as when there are few eligible studies for inclusion in the review.

Consider figure 1, which is adapted from a Cochrane review of 36 psychological interventions for coronary heart disease. The outcome of interest here is reduction in depressive symptoms, for which 11 studies were included. The comparison is any psychological intervention versus control, where control is defined as standard care/treatment as usual (TAU). A fixed-effect meta-analysis was conducted by the authors and a standardised mean reduction of −0.18 (95% CI −0.24 to −0.12) suggests that psychological interventions may affect a modest reduction in depression postcoronary heart disease. However, the p value for the χ² statistic provides extremely strong evidence against the null hypothesis of homogeneity (ie, that interventions are estimating a single underlying treatment effect).

The I² statistic suggests that 75% of the variation between studies is attributable to heterogeneity and not chance. To account appropriately for the observed between-study variation a random-effect meta-analysis may have been more appropriate. However, this would still only answer an ‘in principle’ question of general effectiveness and results would not enable a clinician to select a specific psychological intervention for their patient. The meaningful analysis of complex interventions can therefore pose problems if a ‘lumped’ approach is followed. Further exploration can be achieved by conducting an a priori specified subgroup analysis. Figure 2 shows subgroup analysis by mode of therapy delivery, however this does not appear to explain the observed heterogeneity (individual therapy I²=77%; group therapy I²=86%), and the test for subgroup differences is non-significant (p=0.31). In principle, the interventions could be further subgrouped such as ‘individual+weekly meetings’ or ‘group+weekly meetings+telephone support’. However, caution should be exercised.
since such analyses may suffer from low power due to the small
number of included studies in each grouping. If the purpose of a review
is to investigate which type of psychological intervention is effective, or
which intervention characteristics are effective, then a review which
categorises the intervention characteristics and ‘splits’ the analysis by
intervention type may be the more appropriate and robust strategy. This
can either be achieved as a series of separate reviews or as separate
analyses within the same review.16

CATEGORISATION OF INTERVENTION CHARACTERISTICS

There are a number of ways in which a splitting approach can be
applied for meta-analyses of complex interventions. One possibility is to
use the theoretical underpinning of the interventions to construct ‘clinically
meaningful units’, which should be specified a priori.19 In clinical
psychology this might include classification by intervention modality
such as cognitive–behavioural therapy (CBT), humanistic therapy or
behavioural therapy (BT). In reviews of mental ill-health prevention,
interventions could be grouped by psychological or behavioural theory,
such as the theory of planned behaviour, health belief model, social
cognitive theory and so on.20 21 In figure 3 the psychological interven-
tions for coronary heart disease have been categorised according to
intervention modality. The information was obtained from the
Characteristics of Studies tables included in the original Cochrane
review. The three modalities were CBT, BT and counselling-based inter-
ventions. Of these, we note only BT was associated with a reduction in
depression and the I² is 0%. However, the between-study heterogeneity
is still very high for CBT and counselling, and further investigation is
warranted (note that estimates of heterogeneity become problematic
when few studies are involved). One could disaggregate the intervention
modalities further, for example, under CBT one might be interested in
problem-solving therapies or rational-emotive behavioural therapies.22
Note however that a balance needs to be found between a sufficiently
detailed categorisation that can explain heterogeneity and sufficient
numbers of studies for statistical power and to avoid spurious findings.

---

**Figure 1** Forest plot from a fixed effect meta-analysis comparing “any psychological intervention” vs control (treatment as usual). Outcome is reduction in depressive symptoms. Analysis conducted in Review Manager.43

**Figure 2** Forest plot from a fixed effect meta-analysis comparing “any psychological intervention” vs control, interventions categorised as “individual therapy” and “group therapy”. Outcome is reduction in depressive symptoms. Analysis conducted in Review Manager.43

---

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Std. Mean Difference (IV, Random, 95% CI)</th>
<th>Study or Subgroup</th>
<th>Std. Mean Difference (IV, Random, 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>-0.60 [-1.12, -0.09]</td>
<td>Total (95% CI)</td>
<td>-0.29 [-0.48, -0.10]</td>
</tr>
</tbody>
</table>
| Burgess           | 0.40 [0.06, 0.74]                      | Heterogeneity: Tau² = 0.08; Chi² = 50.32, df = 10 (P < 0.00001); I² = 80%
| Elderen           | -0.21 [-0.75, 0.34]                   | Test for overall effect: Z = 2.98 (P = 0.003)
| ENRICHID          | -0.33 [-0.43, -0.24]                  |                           |
| HoffmanBang       | -0.39 [-0.88, 0.10]                   |                           |
| Johnston          | -0.57 [-1.09, -0.06]                  |                           |
| Jones             | -0.01 [-0.10, 0.07]                   |                           |
| Lewin             | -0.61 [-0.98, -0.23]                  |                           |
| Stern             | -0.30 [-0.63, 0.23]                   |                           |
| Thompson          | -0.39 [-0.92, 0.14]                   |                           |
| Toolbert          | -0.91 [-1.75, -0.08]                  |                           |
| Subtotal (95% CI) | -0.34 [-0.72, 0.04]                   |                           |

**1.1.1 Individual therapy**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Std. Mean Difference (IV, Random, 95% CI)</th>
<th>Study or Subgroup</th>
<th>Std. Mean Difference (IV, Random, 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>-0.60 [-1.12, -0.09]</td>
<td>Total (95% CI)</td>
<td>-0.29 [-0.48, -0.10]</td>
</tr>
</tbody>
</table>
| Burgess           | 0.40 [0.06, 0.74]                      | Heterogeneity: Tau² = 0.17; Chi² = 21.31, df = 5 (P = 0.0007); I² = 77%
| Elderen           | -0.21 [-0.75, 0.34]                   | Test for overall effect: Z = 1.77 (P = 0.08)
| ENRICHID          | -0.33 [-0.43, -0.24]                  |                           |
| HoffmanBang       | -0.39 [-0.88, 0.10]                   |                           |
| Johnston          | -0.57 [-1.09, -0.06]                  |                           |
| Lewin             | -0.61 [-0.98, -0.23]                  |                           |
| Stern             | -0.30 [-0.63, 0.23]                   |                           |
| Thompson          | -0.29 [-0.92, 0.14]                   |                           |
| Toolbert          | -0.91 [-1.75, -0.08]                  |                           |
| Subtotal (95% CI) | -0.34 [-0.72, 0.04]                   |                           |

**1.1.2 Group therapy**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Std. Mean Difference (IV, Random, 95% CI)</th>
<th>Study or Subgroup</th>
<th>Std. Mean Difference (IV, Random, 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>-0.60 [-1.12, -0.09]</td>
<td>Total (95% CI)</td>
<td>-0.29 [-0.48, -0.10]</td>
</tr>
</tbody>
</table>
| Burgess           | 0.40 [0.06, 0.74]                      | Heterogeneity: Tau² = 0.05; Chi² = 27.96, df = 4 (P < 0.0001); I² = 86%
| Elderen           | -0.21 [-0.75, 0.34]                   | Test for overall effect: Z = 1.96 (P = 0.05)
| ENRICHID          | -0.33 [-0.43, -0.24]                  |                           |
| HoffmanBang       | -0.39 [-0.88, 0.10]                   |                           |
| Johnston          | -0.57 [-1.09, -0.06]                  |                           |
| Jones             | -0.01 [-0.10, 0.07]                   |                           |
| Stern             | -0.30 [-0.63, 0.23]                   |                           |
| Toolbert          | -0.91 [-1.75, -0.08]                  |                           |
| Subtotal (95% CI) | -0.25 [-0.49, 0.00]                   |                           |

Heterogeneity: Tau² = 0.05; Chi² = 27.96, df = 4 (P < 0.0001); I² = 86%
Test for overall effect: Z = 1.96 (P = 0.05)
Test for subgroup differences: Chi² = 0.18, df = 1 (P = 0.67), I² = 0%
COMPONENTS-BASED NETWORK META-ANALYSIS

The obvious difficulty for standard, pairwise meta-analyses which seek to disaggregate complex interventions is that there are typically too few studies to allow clinically useful ‘splitting’ and inevitably, some degree of aggregation is needed for a meta-analysis to be conducted. In network meta-analyses (NMAs), however, the potential for disaggregating the intervention is more promising. A NMA is an extension of traditional pair-wise meta-analysis to include multiple interventions, as long as these interventions form a connected network of evidence (see figure 4, which shows a ‘star network’ where all interventions have been compared with the same common comparator—TAU). A key advantage of NMA is that it produces summary estimates of relative effectiveness regardless of whether interventions have been compared directly and ranks them according to the outcome measured (eg, effectiveness or safety). An additional advantage of NMA is that it allows more studies to be combined, as long as they connect to the network (eg, CBT vs counselling studies could be added to figure 4), bringing increased precision in the estimated intervention effects and the potential to explore statistical heterogeneity. For further details on the statistical methodology readers should see refs.23–25 and for a discussion of the implications for systematic review methodology see ref. 26.

In figure 5 the findings from the NMA are reported not only for the comparisons on which there is direct evidence but also for those where it is absent for example, BT versus CBT. There is substantial uncertainty surrounding the pairwise estimates of intervention effect.
and the only comparison that reaches conventional statistical significance is BT vs TAU (note the NMA was performed in a Bayesian framework, which accounts for the wider CIs when compared to figure 3). On the basis of this NMA, CBT is ranked 3rd (95% CIs 1st to 4th) best in terms of reducing depressive symptoms, BT is ranked 1st (95% CIs 1st to 3rd) and counselling is ranked 2nd (95% CIs 1st to 4th). TAU is the ‘worst’ intervention. Note the CIs around the rankings reflect the considerable uncertainty observed in the effect estimates. In NMA a single between-study heterogeneity parameter is typically assumed.23 Here the estimate of $\tau^2$ is 0.11 which might be considered to represent a moderate level of heterogeneity. Note that the NMA assumes that the heterogeneity is the same regardless of which comparison is being made. This may not be appropriate here, since we found more heterogeneity in the CBT vs TAU comparison than for the other comparisons (figure 3). This suggests that the CBT classification may be too broad to capture the complex nature of CBT interventions.

**MULTICOMPONENTS-BASED NMA**

Within an NMA framework, the analyst has greater flexibility to evaluate complex multicomponent interventions and to investigate whether interventions with a particular component(s) are more likely to be effective. Components are defined as the ‘active ingredients’, ‘intervention techniques’ or ‘elements of an intervention that have the potential to causally influence outcomes’.8 As such they may be classified on practical elements, for example, activities, mode of delivery, setting and/or on theoretical underpinnings of the intervention. If there are common components across all interventions in the network, the components effectively become the intervention ‘nodes’ in the network and an NMA can be conducted. Figure 6 represents a multicomponents-based network plot for the coronary heart disease example. Welton et al20 conducted a components-based NMA for the coronary heart disease network. Interventions were classified according to five key components; educational, behavioural, cognitive, relaxation and psychosocial support. Describing their model as a meta-regression-based extension to NMA, three models were evaluated in a Bayesian framework: (1) an additive main effects model which assumes that the effect of each component adds (ie, no synergistic or antagonistic effects), (2) a two-way interaction model (allowing pairs of components to have either a bigger or smaller effect than would be expected from the sum of their effects alone) and (3) a full-interaction model for interventions described as having >2 components (eg, cognitive+behavioral+support). To illustrate, their results for the depression outcome are shown in figure 7 for the main effects additive model. This analysis answers the question “Which intervention component has the greatest probability of being most effective?” Compared to the broader categorisation used in figure 5, having broken down interventions into their component parts heterogeneity is now reduced; $\tau^2=0.03$. There is some evidence that an intervention with a cognitive and/or behavioural component(s) was associated with a reduction in standardised mean

---

**Figure 5** Standardised mean difference estimates from NMA of each psychological intervention versus every other. Interventions analysed as “clinically meaningful units”. Outcome is reduction in depressive symptoms. Analysis conducted using OpenBUGS44 and results plotted using RevMan.43

**Figure 6** Network plot for a multi-component-based network meta-analysis: components identified from a full interaction model for psychological interventions for coronary heart disease. Each node represents a component, or combination of components, identified from the interventions included in the NMA for psychological interventions for coronary heart disease review8 for outcome of depression. The diagram depicts all possible combinations of components from the full interaction model. The solid lines indicate where there was direct information available between component comparisons. TAU/T, treatment as usual; EDU/E, educational; BEH/B, behavioural; COG/C, cognitive; RELAX/R, relaxation; SUP/S, support. + indicates a combination of components, for example, ‘E+B’ is educational and behavioural components.

**Figure 7** Components based NMA: additive main effect model for psychological interventions relative to usual care. Analysis conducted using OpenBUGS44 and results plotted using RevMan.43
depression score; for the cognitive component the pooled Standardised mean difference (SMD) was −0.26 (95% credible interval: −0.55 to 0.02) and for behavioural it was SMD −0.24 (95% credible interval: −0.42 to −0.06).

DISCUSSION

Component-based NMA is an option for the synthesis of complex interventions in the presence of heterogeneity. Of course, intervention categorisation is only one dimension contributing to heterogeneity in meta-analyses of complex interventions. In the above example, heterogeneity was explained by intervention definition but this may not be the case for all examples, where additional factors may cause residual heterogeneity (e.g., an imbalance of effect modifiers across studies). A possible source of confounding here is the control intervention. In RCTs in clinical psychology and psychiatry control interventions may take several forms—waiting list controls are common as are no intervention controls.41 A psychological placebo, where the intervention is regarded as inactive by the researchers but is judged as active by the participants, may be used. Similarly an attention placebo could be used where the control mimics the theoretically inactive elements of an intervention, but not the active elements.32 33 Reviewers of complex interventions should also be mindful that TAU and standard care may differ across settings, contexts and countries, even though systematic reviews have traditionally lumped these into a single control.9 34 Unfortunately, due to the small number of studies in the psychotherapy for coronary heart disease meta-analysis, further disaggregating by control intervention is of questionable value.

Component-based systematic reviews are becoming increasingly common as analysts realise the importance of identifying and investigating heterogeneity, regardless of its inevitability.35 36 However, one difficulty in a components-based approach is the identification of distinct components from the published literature.37 Complex interventions may not be described in sufficient detail to allow dismantling of key ingredients. Recent reporting guideline initiatives, such as CRoDiCI 2 (Criteria for Reporting the Development and Evaluation of Complex Interventions in healthcare: revised guideline)38 and CONSORT-SPI (Social and Psychological Interventions)39 seek to address this. The MRC’s recent guidance on process evaluations may also help identification of components, and assessment of delivery and fidelity of complex interventions.30 How to classify complex interventions and disaggregate the multiple interacting components within them is an area of ongoing interest. Several taxonomies have been developed; some designed for use in specific clinical areas and others are generic.40 Further research is needed to assess the application of taxonomies across clinical areas. Logic models describing the mechanisms of action and causal pathways of interventions are increasingly used to structure systematic reviews of complex interventions42 and could also be used to inform the classification of intervention components. What is clear, however, is that whichever approach the analyst chooses to categorise interventions it is desirable that components be specified a priori, and published in a protocol before data extraction to avoid data-driven decisions and reduce the likelihood of spurious findings.

Funding DMC is supported by a Medical Research Council Population Health Scientist fellowship award G0902118. This work was undertaken with the support of The Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (IDCfPHer), a UKCRC Public Health Research Centre of Excellence. Joint funding (MR/K023233/1) from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, the Welsh Government and the Wellcome Trust, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

Competing interests None declared.

Provenance and peer review Commissioned; externally peer reviewed.

doi:10.1136/eb-2015-102275

Received 15 November 2015; Revised 16 November 2015; Accepted 19 November 2015

REFERENCES


Evid Based Mental Health: first published as 10.1136/eb-2015-102275 on 20 January 2016. Downloaded from http://ebmh.bmj.com/ on February 4, 2021 by guest. Protected by copyright.
44. Available from http://www.openbugs.net/