Conceptual and statistical analysis of complex interventions in the presence of confounding variables: An example from public health

Alison O’Mara-Eves*1, Ginny Brunton1, Sandy Oliver1, Josephine Kavanagh1, Farah Jamal2, James Thomas1

1EPPI-Centre, SSRU, Institute of Education, London; 2Institute for Health and Human Development, University of East London

Background: Meta-analyses of complex interventions are challenging because causality operates through multiple paths and confounding variables can be difficult to distinguish.

Objectives: To meta-analyse public health interventions that engage members of the community in their conception, design, or delivery. To disentangle intervention complexity by analysing according to their theories of change.

Study selection criteria: Published after 1990; outcome or process evaluation; community engagement intervention; written in English; reported health or community outcomes; study populations or differential impacts reported according to social determinants of health.

Analysis: Intervention complexity was examined by conceptualising, operationalising, and mapping their theories of change; and through random effects subgroup analyses.

Main results: 131 studies were included in the synthesis. Three main theories of change were identified, which were useful in describing trends in intervention effectiveness. Statistically significant between-group differences were not detected, since there were likely to have been too many confounding variables.

Conclusions: Intervention complexity in systematic reviews can be addressed through examining theories of change and trends in effect size estimates. Such complexity appears to defy current meta-analytical methods when confounding variables undermine analysis of variance.

*Corresponding author: Dr. Alison O’Mara-Eves, SSRU, Institute of Education, London, WC1H 0NR, United Kingdom. E-mail: a.omara-eves@ioe.ac.uk.
Background

Complex interventions are widely used in a range of social policy areas. The UK Medical Research Council guidance on complex interventions suggests that the following characteristics can make an intervention complex (1):

- “Number of interacting components within the experimental and control interventions
- Number and difficulty of behaviours required by those delivering or receiving the intervention
- Number of groups or organisational levels targeted by the intervention
- Number and variability of outcomes
- Degree of flexibility or tailoring of the intervention permitted” (p. a1655).

Craig et al. (1) argue that a theoretical understanding of the underpinning causal mechanisms is needed when evaluating complex interventions. This is because the cause and effect relationship can be obscured by complex factors such as those listed above, making it difficult to disentangle the actual effect of the intervention on a given outcome. By specifying the particular causal mechanism underpinning a proposed effect, we can attempt to isolate that pathway from the potential confounders introduced through complexity.

Complex interventions pose particular problems for systematic reviewing. Different interventions often posit (and test) different mechanisms or pathways through which an intervention may affect an outcome variable. Differences in the observed effect of an intervention (beyond random chance) might therefore be attributable to the complexities listed above or to differences in the causal mechanisms underpinning the interventions. This raises questions about the comparability of such interventions and whether they should be combined in a synthesis.

Whilst in some cases it might not be appropriate to combine studies with different proposed causal pathways, there are other instances in which it might be particularly interesting to explore the relative effectiveness of different causal mechanisms. In this paper, we present an example of a systematic review from public health in which we
explored causal mechanisms as a source of complexity because there are a number of competing ideas about how this particular intervention (community engagement) could lead to improved outcomes.

**A public health example: community engagement**

Historically, interventions and actions to promote health were driven by professionals with little or no input from the targeted populations (2). A more recent movement from practitioners, policymakers, and researchers has been to engage members of the community in public health interventions (e.g., 3, 4). Community engagement has been broadly defined as “involving communities in decision-making and in the planning, design, governance and delivery of services” ((5) p 11). Community engagement activities can take many forms and are usually described in terms of five levels of engagement (from least to most engaged): information-giving, consultation, joint decision-making, acting together, and supporting independent community interests (6).

Previous reviews of the literature on community engagement in public health interventions suggest potential social improvements (e.g., improved social capital and neighbourhood cohesion) but report unclear effects on morbidity, mortality and health inequalities (7, 8), largely due to a lack of identified evaluations. To explore this issue further, our team conducted a mixed methods systematic review to examine whether community engagement strategies in public health had an impact on health outcomes (9). We took a broader approach to the literature than previous reviews to ensure that a large number of evaluations were identified.

This paper presents components of the results of a theoretical synthesis and a statistical meta-analysis reported in the full report (9), to demonstrate an approach to exploring causal mechanisms in complex interventions. The original project (9) was a multi-method systematic review containing five components: a map of the evaluative and theoretical literature that describes the scale and range of community engagement interventions; a meta-analysis of evaluations; a thematic summary of process evaluations linked to evaluation studies focused on health inequality policy priority areas; an analysis of costs and resources; and a newly developed conceptual framework that brings together the learning from all components of the project.
Methods

The complete and detailed methods of the full project are reported by O’Mara-Eves and colleagues (9). Here, we summarise the details pertinent to this paper.

Aims

A key aim of the research presented here was to meta-analyse public health interventions that engage members of the community in their conception, design, or delivery, specifically those targeted at disadvantaged groups. We further sought to disentangle intervention complexity by analysing the interventions according to their theories of change. A theory of change describes the mechanisms through which an intervention is proposed to lead to an outcome. In order to meet this aim, we first needed to identify the theories of change evident in the community engagement literature. As such, an additional aim was to conduct a theoretical synthesis to identify theories of change, and then operationalise the theories for inclusion in the statistical analysis.

The aim of this paper is to present an approach to examining conceptual complexity in a systematic review of complex interventions through the examination of theories of change identified in the evidence base.

The theoretical synthesis

The theoretical synthesis was similar in some respects to Pawson’s work on realist synthesis (10) and examined the theories, mechanisms and contexts of community engagement. It did not, however, attempt to engage in causal reasoning, leaving this task to the meta-analysis.

The theory-building nature of this synthesis led the research team to use methods of study identification and synthesis more appropriate for conceptual analysis research synthesis. In this type of synthesis, searching aims to build an understanding of a particular phenomenon by gathering a number of articles that present different perspectives on that phenomenon. Once a sufficient range of ideas have been identified, studies which do not add anything new to the topic are put to one side: in
effect, a saturation of perspectives has been reached (11). We then extracted information about theories, mechanisms and contexts of community engagement, to identify any distinct theories of change.

**Study identification and selection for the meta-analysis**

A sample of the search syntax used is presented in Appendix A and the detailed screening and inclusion criteria are recorded in the full project report (9); here we briefly summarise the process. We searched the following sources without language restriction for systematic reviews (SRs) of public health interventions: Cochrane CDSR and CCTR, Campbell Library, DARE, NIHR HTA programme website, HTA database, and DoPHER. Through the identified SRs, we collated a database of primary studies that appeared to be relevant, and screened the full-text documents of those primary studies against our inclusion criteria. In parallel, we searched the NHS EED and TRoPHI databases for additional primary studies. We also contacted key authors and citation chasing of linked studies to identify further studies.

Full-text reports of all systematic reviews on public health topics identified through these sources were retrieved; their summary tables were then scanned to locate relevant trials. A secondary screening of titles and abstracts eliminated studies published before 1990 and from non-OECD countries. All full-text reports of relevant trials were subsequently retrieved, screened and included if they:

- Reported primary research;
- Were not a Masters thesis;
- Included intervention outcome and/or process evaluations;
- Focused on community engagement as the main approach;
- Contained a control or comparison group;
- Characterised study populations/reported differential impacts of social determinants of health according to: place of residence, race/ethnicity, occupation, gender, religion, education, socio-economic position, and social capital, plus other variables describing ways in which people may be systematically disadvantaged by discrimination (including sexual
orientation, disability, social exclusion, and challenging life transitions such as teenage pregnancy): the ‘PROGRESS-Plus’ framework (12); and

- Reported health or health-related (including cost) effectiveness outcomes and/or process data.

Due to the large number of studies identified for inclusion in the review of community engagement interventions (n = 319; see full report for details), and in order to align our work with current UK health policy priority areas, we narrowed the scope of health topics included in the meta-analysis by focusing on the policy objective areas identified in the Marmot Review of health inequalities, ‘Fair Society, Healthy Lives’ (13), plus the key modifiable health risks identified in the Marmot Review (smoking, alcohol abuse, substance abuse, and obesity). This led to a final sample of 131 studies.

**Data extraction**

Data were extracted from the 131 studies on models, approaches, and mechanisms of community engagement; health topic; participant characteristics; geographical and contextual details; costs and resources; and processes.

Effect size estimates were calculated using standard techniques (14), adjusting for cluster allocation (15) where necessary. Effect size estimates based on continuous data were calculated as the standardised mean difference (represented by $d$), while logged odds ratios were used for binary outcomes. Logged odds ratios were transformed to standardised mean differences using the methods described by Chinn (16), so that the different types of effect size estimates could be included in the same analyses$^1$. A positive $d$ indicates that participants in the treatment group, on average, scored higher than those in the control group. An effect size estimate of $d = 1.0$ means that participants in the treatment group scored – on average – one standard deviation higher than the control group on the particular outcome measure.

$^1$ We conducted a sensitivity analysis to test whether $d$ effect size estimates based on binary data were statistically similar to $d$ effect size estimates based on continuous data. Although pooled binary outcomes tended to be slightly smaller than pooled continuous outcomes, this difference was not statistically different ($Q_b (1) = 3.03, p = .08$).
For the purposes of this report, we focus on the results for health behaviour outcomes². Outcomes included were: alcohol abuse, antenatal (prenatal) care, breastfeeding, cardiovascular disease, child illness and ill health, drug abuse, healthy eating, immunisation, injury / safety, parenting, physical activity, smoking cessation, and smoking/tobacco prevention.

Data analysis
We conducted random effects model analyses (ANOVAs and multiple regressions) with maximum likelihood estimators, following the methods described in (14). We used SPSS macros written by David Wilson³ to run the models. For the homogeneity analyses, between groups $Q$-statistic ($Q_B$) indicates the extent to which the categories of studies differ from each other; and within groups $Q$-statistic ($Q_W$) indicates the extent to which the effect size estimates within a category differ from each other.

In this paper, we only report the findings related to the analysis of theories of change: the focus of the current paper is on demonstrating how theories of change can be examined systematic reviews of complex interventions. In the full report, we conducted an assessment of the risk of publication bias. Additionally, the following variables were included in subgroup or sensitivity analyses:

- Theory of change underpinning the intervention
- Single or multiple components to the intervention
- Health topic
- Universal versus targeted approach
- Intervention setting
- Intervention strategy
- Intervention deliverer
- Duration of the intervention
- PROGRESS-Plus group

² In the full report, we further report results for health states (e.g., mortality, morbidity), participant self-efficacy pertaining to the health behaviours, participant social support in relation to the health behaviours, community outcomes, and outcomes for the community members involved in the intervention. These are not reported here for simplicity – they are not needed to demonstrate the key point about modelling theories of change to explore complexity.
³ http://mason.gmu.edu/~dwilsonb/ma.html
We chose to focus our reporting of the results on the trends in pooled effect size estimates, rather than between-group statistical significance. Typical meta-analyses attempt to infer findings from the sample to a hypothetical population. This approach would be problematic for this review because the issues that were explored—community engagement and health inequalities—are so broad and difficult to define that it is impossible to know exactly to what population the results of any inferential statistics would apply. Instead, we emphasised observed trends, to help disentangle some of the differences between the types of evidence we have collected. This can help us to understand what might occur in other similar studies not included in the review, but not in any one specific situation because the causal pathways are complex and potentially unique to each study.

**Results**

Intervention complexity was examined by conceptualising, operationalising, and mapping the interventions’ theories of change. The theories of change were then included as a moderator variable in random effects subgroup analyses. The results of these analyses follow.

**Identifying and operationalising theories of change**

Three overarching conceptualisations of engagement emerged from the theoretical synthesis, which combined aspects of purpose, theory, and the way that many interventions occur in practice.

The first model is centred on the concept of **empowerment**: the idea that change is facilitated where the health need is identified by the community and they determine their own courses of action (17). This is believed to be effective because democratic input can encourage community members to feel that the intervention is worthwhile,
that improvements are possible, and that the intervention suits their needs—all of which in turn can improve commitment to the intervention and the appropriateness and acceptability of the intervention to the community. Empowerment interventions were operationalised as studies in which the community was explicitly involved in identifying the health problem/need. This classification was regardless of whether community members were involved in the design and/or delivery of the intervention, because the main distinguishing feature of empowerment models was that the community mobilised themselves into action.

In the second model, the need for intervention is usually identified by observation that is external to the community, but the views of stakeholders are sought with the belief that the intervention will be more appropriate to the participants’ needs as a result. We identified two main mechanisms through which stakeholder views are sought in the design or planning of the intervention: through collaboration with the community, or through consultation with the community. This second group of interventions were operationalised as studies in which the community collaborated in or were consulted about the design or planning of the intervention but were not involved in identifying the health need; they may or may not have been involved in the delivery of the intervention. ‘Collaboration’ and ‘consultation’ theories of change were treated as separate groups in the analysis because of the differences in the relationships and power relations inherent in these two approaches.

The final model is concerned with traditional models of peer- or lay- delivered interventions. In these interventions, the community ‘ingredient’ is included through the use of specific people in delivering the intervention; however, beyond informing peers about the intervention, no peer involvement in the intervention’s planning or delivery occurs. In this model, change is believed to be facilitated by the credibility, expertise, or empathy that the community member can bring to the delivery of the intervention. Traditional models of peer- or lay- delivered interventions were operationalised as studies in which community members led or collaborated in the delivery of the intervention but were not involved in the design of the intervention, nor were they involved in identifying the health need.
Description of the studies included in the meta-analysis

Of the 131 studies included in the meta-analysis, 113 (86.3%) were conducted in the USA, five (3.8%) were conducted in the UK, five (3.8%) were conducted in Canada, and eight (6.1%) were conducted in other OECD countries. In terms of publication date, 63 (48.1%) were published in the 1990s, 62 (47.3%) in the 2000s, and 6 (4.6%) in 2010 or later.

The interventions were conducted over a range of health topics and settings. The most commonly-targeted health issue was substance abuse (n = 18, 13.7%), followed by cardiovascular disease (n = 14, 10.7%). The most common setting for interventions was in the community (n = 56, 42.7%). Many interventions were also conducted via media tailored to the participants’ needs (e.g., tailored newsletters or information sheets, n = 53, 40.5%) or mass media (n = 21; 16%); such interventions could be delivered wherever the participant was located.

Most of the interventions included multiple intervention strategies. The most common strategy was education provision (n = 105, 80.2%). Advice (n = 71, 54.2%), social support (n = 58, 44.3%), and skill development training (n = 51, 38.9%) were also common strategies. Interventions were most commonly delivered by non-peer community members (n = 58, 44.3%) and peers (n = 49, 37.4%).

Sixty-nine (52.7%) trials were considered to have an overall low risk of bias and 62 (47.3%) trials were considered to have an overall high risk of bias.

Overall pooled effect size estimate

Although there were 131 studies included in the broader review (9), only 105 reported health behaviour outcomes and were included in the following analyses.

The pooled effect size estimate across interventions is positive (indicating that the outcomes measured were in favour of the treatment group) and statistically significant (as indicated by the p-values and 95% confidence intervals) for these four outcomes. For health behaviours, the pooled effect size estimate \( d = .33 \) (95% CI .26 to 40). The statistical significance of the pooled effect means that the effect size estimate is
significantly different from a null effect in which there are no differences between the intervention group and the comparison group.

There was, however, significant heterogeneity in the studies ($\tau^2 = .093; \Gamma^2 = 82.8\%; Q(104) = 604.62, p < .001$). As such, we conducted moderator analyses to attempt to explain variation in the observed effectiveness of the interventions. Critical to this paper’s aims, we explored the variation between studies grouped by the theories of change identified in the theoretical synthesis.

**Analysis by theories of change**

We ran an analysis to compare the effectiveness of interventions employing one of four different operationalisations of the theories of change on health behaviour outcomes. Although there was no overall significant difference between the studies grouped by theory of change, some clear trends emerge (see Table 1). Interventions that engaged the community through peer delivery of the intervention had the largest pooled effect size estimate, while interventions that adopted self-mobilisation, design collaboration, or design consultation theories of change (whether implicitly or explicitly) had overall effect size estimates that were similar in magnitude to one another but substantially lower than lay-delivered interventions. Interventions that did not fit into one of the four main theories of change categories had the smallest pooled effect size estimate.

We conducted supplementary analyses to try to explain why the lay-delivered interventions might be more effective. One explanation that we considered was the size of the study. Based on our observations of the studies, we suspected that the lay-delivered interventions tended to be smaller-scale and usually more likely to be one-on-one, compared to interventions where the community was involved in the design of the intervention. From Table 1, we can see that the mean sample size for studies that only involved the community in the delivery of the intervention is much smaller than in studies based on alternative theories of change. Post hoc analyses of a one-way ANOVA with sample size as the dependent variable and the different theories of change as the factors indicate that the mean sample size for the lay-delivered interventions is statistically significantly smaller than for the interventions in which the community identified the health need.
Discussion

Overall, community engagement interventions are effective in improving health behaviours. However, there was significant heterogeneity (both statistical and conceptual), which made this synthesis of complex interventions especially challenging.

In an attempt to disentangle some of the conceptual complexity and to explain some of the statistical heterogeneity, we analysed the studies according to their implicit or explicit theories of change. The theories of change had been identified through a theoretical synthesis.

We compared the effectiveness of interventions based on four different theories of change in the synthesis of effectiveness data. The results suggest that lay-delivered interventions tend to have larger effects than interventions based on empowerment or patient/consumer involvement, although this trend did not significantly explain statistical variation in the observed effect across studies. We propose that this association is likely to be confounded with other factors, such as intervention intensity and exposure (lay-delivered tend to be more intense, one-on-one or small group interventions, than other intervention types). For such models, we might expect to see large effects over a narrow range of outcomes, as opposed to empowerment models that might have smaller effects over a broader range of health and social outcomes. Unfortunately, there were insufficient data to test these relations adequately.

Issues in interpreting statistical findings in reviews of complex interventions
Significant statistical heterogeneity was expected in this review, and indeed the exploration of this heterogeneity was part of its design. When operating across such a wide range of topics, populations and intervention approaches, however, there is a disjunction between the conceptual heterogeneity implied by asking broad questions and the methods for analysing statistical variance that are in our ‘toolbox’ for answering them.
Analysing the variance ‘explained’ by specific sub-groups of studies according to our conceptual framework rarely reached accepted standards for statistical significance. This is inevitable however, because conceptual homogeneity was never achieved through such a sub-division: each type of approach to engagement was observed across populations, topics, outcomes and a wide range of other unknown variables; we would therefore never reach the position of being able to say that the studies within a given sub-group differed only due to sampling error / variance. (Or that any of our sub-divisions was the only way of partitioning the studies present.) In other words, potential confounding variables or interactions amongst variables made it difficult to disentangle unique sources of variance across the studies.

In the context of our analysis this debate is relevant because statistical tests for significance are unlikely to yield statistically significant findings due to complex heterogeneity in the dataset. We are therefore left with an interpretive challenge: do we adhere strictly to the $p > 0.05$ convention before accepting that a given sub-group analysis is meaningful; or do we place more importance on the magnitude of the differences of effect size estimates between sub-groups? In this review we have attempted to plot a path somewhere between the two extremes. We have tested and reported statistical significance, but have also drawn tentative conclusions from the directions and magnitudes of effects whether or not standard statistical significance had been achieved.

**Recommendations for systematic reviews of complex interventions**

Intervention complexity in systematic reviews can be addressed through examining theories of change and trends in effect size estimates. By examining theories of change, we can:

- (a) Determine the appropriateness of combining interventions with different causal mechanisms,
- (b) Articulate the underlying assumptions that are being tested in both the evaluations and in the review,
- (c) Explore conceptual complexity across studies, and
- (d) Attempt to differentiate between statistical variance due to the effects of the intervention and those due to other features of complexity.
At the moment, such complexity appears to defy current meta-analytical methods when confounding variables undermine analysis of variance. In many meta-analyses, it is not possible to include different types of complexity (e.g., conceptual and those listed in the MRC guidance at the start of this paper) because of limitations to meta-analytic modelling. One of the most common limiting factors is a lack of data: a large number of effect sizes are needed to test for simultaneous explanatory variables. Alternatives to meta-analysis are emerging (e.g., qualitative comparative analysis (18)) that appear to be well-equipped to deal with multi-complex datasets and could be employed in the context of systematic reviews.

**Acknowledgements**

David McDaid, Tihana Matosevic, and Angela Harden, for their contributions to the full project, and Katherine Twamley, Irene Kwan, Carol Vigurs, and Jenny Woodman who helped with screening and/or document retrieval.

**Funding**

This report refers to independent research commissioned by the National Institute for Health Research (NIHR). Any views and opinions expressed herein are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the PHR programme, or the Department of Health.

**References**


Table 1 - Results of the random effects ANOVA analyses by theory of change for health behaviour outcomes

<table>
<thead>
<tr>
<th>Theory of change</th>
<th>Pooled ES estimate</th>
<th>95% CI</th>
<th>n</th>
<th>Average sample size (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empowerment: Community identified health need</td>
<td>.31***</td>
<td>.14, .48</td>
<td>17</td>
<td>1067 (226.30)</td>
</tr>
<tr>
<td>Collaboration to design more appropriate intervention</td>
<td>.32**</td>
<td>.13, .51</td>
<td>16</td>
<td>1924.91 (910.74)</td>
</tr>
<tr>
<td>Consulted to design more appropriate intervention</td>
<td>.25***</td>
<td>.12, .38</td>
<td>27</td>
<td>848.67 (184.53)</td>
</tr>
<tr>
<td>Lay-delivered to enhance credibility, expertise, or empathy</td>
<td>.47***</td>
<td>.34, .60</td>
<td>38</td>
<td>309.74 (48.21)</td>
</tr>
<tr>
<td>Other</td>
<td>.17</td>
<td>-.07, .42</td>
<td>7</td>
<td>757.14 (213.08)</td>
</tr>
</tbody>
</table>

** p < .01, *** p < .001. Statistical significance indicates the effect size estimate is significantly different from zero. Note. ES = effect size estimate, 95% CI = 95% confidence interval of the pooled effect size estimate; n = the number of effect size estimates in the subgroup; SD = standard deviation. Heterogeneity statistics for the meta-analysis: Q_B (4) = 7.80, p = .10; Q_W (100) = 97.63, p = .54.
Appendix A: Sample search syntax

Search syntax used in the Cochrane Databases (searched on 17/8/2011)

- Cochrane Database of Systematic Reviews (Cochrane Reviews)
- Database of Abstracts of Reviews of Effects (Other Reviews)
- Health Technology Assessment Database (Technology Assessments)
- NHS Economic Evaluation Database (Economic Evaluations)

“disadvantage” OR “disparities” OR “disparity” OR “equality” OR “equity” OR “gap” OR “gaps” OR “gradient” OR “gradients” OR “health determinants” OR “health education” OR “health inequalities” OR “health promotion” OR “healthy people programs” OR “inequalities” OR “inequality” OR “inequities” OR “inequity” OR “preventive health service” OR “preventive medicine” OR “primary prevention” OR “public health” OR “social medicine” OR “unequal” OR “variation” AND

“change agent” OR “citizen” OR “community” OR “champion” OR “collaborator” OR “disadvantaged” OR “lay community” OR “lay people” OR “lay person” OR “member” OR “minority” OR “participant” OR “patient” OR “peer” OR “public” OR “representative” OR “resident” OR “service user” OR “stakeholder” OR “user” OR “volunteer” OR “vulnerable” AND

“capacity building” OR “coalition” OR “collaboration” OR “committee” OR “compact” OR “control” OR “co-production” OR “councils” OR “delegated power” OR “democratic renewal” OR “development” OR “empowerment” OR “engagement” OR “forum” OR “governance” OR “health promotion” OR “initiative” OR “integrated local development programme” OR “intervention guidance” OR “involvement” OR “juries” OR “local area agreement” OR “local governance” OR “local involvement networks” OR “local strategic partnership” OR “mobilisation” OR “mobilization” OR “neighbourhood committee” OR “neighbourhood managers” OR “neighbourhood renewal” OR “neighbourhood wardens” OR “networks” OR “organisation” OR “panels” OR “participation” OR “participation compact” OR “participatory action” OR “partnerships” OR “pathways” OR “priority setting” OR “public engagement” OR “public health” OR “rapid participatory assessment” OR “regeneration” OR “relations” OR “support”