Author’s response to reviews

Title: Using Qualitative Comparative Analysis (QCA) in systematic reviews of complex interventions: a worked example

Authors:

James M Thomas (j.thomas@ioe.ac.uk)
Alison O’Mara-Eves (a.omara-eves@ioe.ac.uk)
Ginny Brunton (g.brunton@ioe.ac.uk)

Version: 2
Date: 17 May 2014

Author’s response to reviews: see over
Using Qualitative Comparative Analysis (QCA) in systematic reviews of complex interventions: a worked example: responses to peer review feedback

We would like to thank the peer reviewers for their thoughtful and helpful feedback, and for their supportive comments about the paper as a whole. We respond to each comment in turn below and are grateful for their assistance in identifying areas of the paper where our explanation needed more clarity. We acknowledge that the paper is quite lengthy, and appreciate the time this must have taken.

Reviewer 1: Andrew Booth

Major Compulsory Revisions

None

Minor Essential Revisions

1. Abstract: “the result of a QCA can take the form of a number of configurations of various participant, intervention, and contextual characteristics which are (or are not) present when the intervention has been successful (or not) in obtaining the desired outcome”. While this sentence captures the variation being tackled within the method it may be clearer by breaking this into two sentences one about the presence or absence of characteristics and one about the success or otherwise of the intervention. In addition the use of brackets overtly interrupts the flow of the sentence so commas may be more elegant stylistically.

*We have simplified the sentence to read: “Developed originally in the area of political science and historical sociology, a QCA aims to identify those configurations of participant, intervention, and contextual characteristics which may be associated with a given outcome.”*

2. The “initial theories of change” only appear in the Results paragraph of the Abstract. It might be helpful to introduce them at least briefly in the description of the Methods within the Abstract. I.e. “Having identified initial theories of change we……”

*We have amended the first sentence to read: “Having identified initial theories of change in a previous analysis, we explore the potential of Qualitative Comparative Analysis (QCA) to assist with complex syntheses through a worked example.”*

3. “a basket of possibilities” – this metaphor may be unclear especially to non-English readers. As it is not an international standard unit I would recommend an expression that is less colourful but more literal.

*We have replaced “basket” with “range”.

4. “as it enables the reviewer to compare the relative effectiveness of different interventions, even if there are no studies which compare them directly” – this makes perfect sense to a reader familiar with network MA. However this might be expressed more clearly “as it enables a reviewer to
indirectly compare the relative effectiveness of Intervention A with Intervention C, even where existing studies separately compare Interventions A and C directly with a third Intervention (B)"

*We have changed the text to read: “Network meta-analysis has been receiving widespread interest as it enables a reviewer indirectly to compare the relative effectiveness of Intervention A with Intervention C, even where existing studies separately compare Interventions A and C directly with a third Intervention (B)”*

5. “many areas of public policy, where complex multi-component interventions are tailored for use in different situations. Here, intervention replication is rare” – It would be helpful to reference the rarity of intervention replication within public policy, particularly given most readers will be more familiar with health/public health review topics.

*We have added a reference to Milat et al here.*

6. Because “training” is both an adjective and a verb I would suggest for this particular example that you refer to “training of intervention providers” to avoid the possible ambiguity that these are providers of a training intervention.

*We have changed the text to read: “For example, in an analysis which is examining whether the training of intervention providers results in better outcomes, a correlational analysis will require that the training of intervention providers is associated with good outcomes AND that the absence of their training is associated with poorer outcomes. If there are multiple approaches to achieving effectiveness though, it may be that some (or all) of the interventions where training did not occur had good reasons for this (e.g. they had recruited more experienced providers) but this would not be picked up in the analysis, and the importance of training in the interventions which did train providers would be lost.”*

7. “a common logic discussed in the literature about causation, necessity and sufficiency” – requires at least one reference not necessarily to show it is common but, at least, that it is discussed in the literature!

*We have added references to Parascandola and Weed (2001); and Phillips and Goodman (2006).*

8. Are these “reviewer messages” or “review messages”? – I tend to think of the messages about the quality of the included studies or the research implications as reviewer (i.e. interpretive) messages but what the review tells us about the interventions as “review messages” i.e. they are less moderated/interpreted.

*We have changed the text at the first “messages” to read: “The identification of critical intervention components relates to a common logic discussed in the literature about causation, necessity and sufficiency, and it is a valuable framework to use when thinking about expressing review findings for potential review users...”*
At the second “messages”, the text has been changed to: “When dealing with complex interventions which differ from one another in subtle (and not so subtle) ways, existing methods appear to lack the analytical purchase necessary to generate actionable findings…”

The third mention of “messages” have left as it: “A QCA synthesis may stop at this point, or it may go on to develop theory to explain its findings and to increase generalizable messages.”

9. “A parallel with the synthesis of evaluations of complex interventions is clear here, where we often have a limited number of studies, and a large number of possible factors that might explain differences in their findings (e.g. participants, interventions, context, outcome measurement, study design, comparator, etc...)” This reads like a sentence fragment albeit a lengthy one. “We might draw a clear parallel here with....etc” might read more fluently.

We might compare here with the synthesis of evaluations of complex interventions, where we often have a limited number of studies, and a large number of possible factors that might explain differences in their findings (e.g. participants, interventions, context, outcome measurement, study design, comparator, etc...).

10. “part of the Boolean set logic” You use this expression here but then describe Boolean set logic in the following paragraph. It would be helpful to describe the application after the description, not before.

We have shortened the sentence to read: “In the former, investigations of combinations of conditions are hampered by collinearity (when variables have a linear relationship), whereas the fact that some conditions may be related to one another is expected in a QCA”

11. “Some have suggested that interventions amongst disadvantaged populations that are completely empowerment-driven will be most effective [9, 10], while others have noted that pragmatic interventions that utilise public involvement to improve them and thus make them more effective.[11]” This sentence seems to have one or more words missing in the latter half.

We have amended the sentence to read: “Some have suggested that interventions amongst disadvantaged populations that are completely empowerment-driven will be most effective [9, 10]; while others have noted that pragmatic interventions that utilise public involvement to improve them can also be successful.[11]”

12. Delete first “is” in: “Membership in the conditions is in this dataset is binary”

We have deleted as suggested.

13. “varies between the very cautious “subgroups kill people”, and the pragmatic “...and lack of subgroup analysis kills people”. [23]” – this is very difficult for a reader who does not have sight of the cited paper to understand. A couple of “i.e.”s would help here.
We have amended the text to read: “Standard guidance on this, and views on the type of knowledge generated, varies between the very cautious “subgroups kill people”, and the pragmatic “…and lack of subgroup analysis kills people” (i.e. subgroups may differ from one another due to chance, and so be misleading; but subgroup analysis may be the only way to answer some research questions, such as the impact of social class).”

14. In a couple of places you offer that QCA may be an “alternative (or adjunct)” - as these are not interchangeable it would be helpful to briefly summarise under what circumstances these are most likely to be an “adjunct” and under what circumstances these are most likely to represent a genuine “alternative”.

We have re-written that section of the conclusion to read: “Current statistical methods of synthesis operate well for homogenous datasets, but poorly where there are few replications and where interventions are complex. QCA is a promising method that should be considered when quantitative synthesis cannot explain the between-study heterogeneity observed; in these situations, it might usefully replace the standard fall-back of a narrative synthesis, and suggest ways in which particular combinations of intervention characteristics might be associated with improved outcomes.”

15. The NIHR should be acknowledged as the source of funding for the wider review.

We now acknowledge the NIHR.

16. The typography of the references appears uneven (font size?)

This may have been a result of the pdf conversion process; the fonts are reporting that they are the same size in Word.

Discretionary Revisions

16. “The positive way in which heterogeneity is viewed is also notable; something that is often portrayed as a problem in systematic reviews” – This is an interesting observation which might be slightly expanded to explain that Heterogeneity is typically seen as complicating subsequent explanation whereas in this case it may be viewed as facilitating potential explanation.

We have expanded the text to read: “The positive way in which heterogeneity is viewed is also notable, since the analysis depends upon differences in order to provide explanations; this is something that is often portrayed as a problem in systematic reviews, where heterogeneity can complicate a meta-analysis by suggesting that all studies are not estimating the same underlying effect.”
Reviewer: Ruth Gwernan-Jones

Reviewer’s report:

I have made a small number (12) of comments in the uploaded pdf (I have highlighted the relevant text, with comments written in the box in the margins).

One of these is a major compulsory revision;

10. The definition of the ‘high quality’ case seems to be a weakness in your argument for QCA, because the category actually encompasses a number of separate issues lumped together, and so QCA was incapable of identifying ‘those components of an intervention which are critical to its success’ – I think you need to address and discuss this issue.

We have moved the text which was in a footnote into the main body of the paper: “In this example, these attributes have been combined into a single “quality” intervention characteristic. It would be possible to have a separate condition for each attribute, though one might quickly run into the problems of “limited diversity” identified earlier (please see discussion on ‘compound’ conditions).”

We have also added a new section to the discussion:

SINGLE AND ‘COMPOUND’ CONDITIONS

When QCA is used in a systematic review, the use of ‘compound’ conditions may be required more than in primary research, since we are limited to the information available in publications (and cannot ‘observe’ more about each case, which may be possible in primary research). This means that a balance will need to be struck between parsimony – the simplest possible ‘solution’ and complexity – looking at the fine detail as to how each case may differ from another. Our conditions named ‘intensity’ and ‘quality’ were ‘compound’ conditions to differing degrees: both might be understood to be made up of multiple intervention characteristics. High quality interventions were those which, in line with some of the principles of engagement we had previously identified, appeared to have high quality interactions between intervention deliverer and participant. These could be operationalised in different ways, such as an emphasis on personal contact, multiple support options, tailored to individuals, and delivered in a location which suited women. While the ‘intensity’ condition might appear more homogenous, the details of how ‘intensity’ was operationalised across the interventions differed subtly from one another (e.g. in terms of who was delivering the intervention, where it was delivered, and precisely how frequently). We could have broken down each condition further in order to represent this heterogeneity in detail; but we would simply have ended up with a list of different interventions. Instead, we chose to ‘drive’ our operationalization of conditions according to theory: ‘quality’ was about how well attuned the intervention was to the way that each participant wanted to be ‘engaged’; and ‘intensity’ was about understanding that it was important to institute breastfeeding promptly after birth, before other routines had become established. Focusing on the theory, the “why” a given issue might affect the outcome is a logical way of creating “compound” conditions, since the conditions grouped here will logically co-occur and there may be little to be gained in separating them. A balance needs to be struck however, between the use of compound conditions to reduce the likelihood of running into problems of limited diversity, and the use of conditions that are fine-grained enough to identify causally important differences between interventions.
Ten are minor essential revisions (1-9, 12).

1. Though the worked example involves only studies that include quantitative measurement of intervention effect; is the 'transcendence' related to a means to involve aspects of context and participant in synthesis of quantitative studies, or can QCA be applied to qualitative studies? It would be useful to clarify this.

   *We clarify this now on (new) page 4: “Sets and set relations, Ragin argues, are the basis of almost all social science theory, and it is through the use of set theoretic principles that Ragin seeks to transcend the qualitative / quantitative divide.[4]”*

3. How would it be decided how much the effect size should be altered in relation to the meaning of an outcome? This seems to be a potential strength of QCA over meta-analysis. Though I understand the effect sizes were appropriately scaled, it was not clear to me whether they were 'qualitatively anchored' or not -- this process could be more clearly described.

   *We have referred readers to two chapters in Ragin’s book on calibration by inserting: “Calibration of fuzzy sets is a complex topic, and for further information we recommend Part II of [4].”*

4. This may not be the best phrase as it might be understood to be dismissive of potential complexities and uncertainties

   *We agree and have rephrased as: “Analysing studies in these terms facilitates the identification of necessary and sufficient conditions for the outcome to be obtained; a feature of QCA which may facilitate the translation of its findings to practical situations.”*

5. 'lay-defined' would be a better label for this group (or another that was more descriptive) -- 1-3 can all be understood as approaches to 'empowerment'

   *As these were the labels used in the original report, we would like to retain them in this paper for consistency.*

6. This phrase isn't right -- should the second 'that' be omitted?

   *This text has been rephrased: “Some have suggested that interventions amongst disadvantaged populations that are completely empowerment-driven will be most effective [9, 10]; while others have noted that pragmatic interventions that utilise public involvement to improve them can also be successful.[11]”*

7. It would be useful here to clarify how the 12 studies reported in this study differ from the textbook selection of cases.

   *In the section on outcome calibration we now state: “Note that all but one of the interventions were effective (in a systematic review, it is not always possible to have clearly differentiated positive and
negative cases), so the outcome for the QCA analyses was membership in the set of highly effective interventions.”

The other text we have added in this revision should also help clarify this issue: “In the context of a systematic review however, such purposive (or ‘theoretical’) sampling can be difficult to achieve, since the set of studies available may not contain both positive and negative findings relating to the same outcome.”

8. A little repetitive with the same section given earlier - can they be combined, or simply listed in one of the sections?

We have moved the lower definitions up, and referred the reader to the previous definition, rather than repeat them again here.

9. Should this be 1 or 0?

This should be 0.333, as that was what was used in the analyses. We rated it as “quite low” and treated it as a fuzzy set. It was the only study that we didn’t end up categorising at the extreme ends (“low” or “high”).

12. This response seems a little confrontational – Pawson argues for a generativist stance and against other stances, but he also argues that different kinds of knowledge should be included in a generativist synthesis so I would have thought you are in agreement.

The intention wasn’t to be confrontational, but rather to reflect Ray’s views, which I’m sure would be against specifying the use of a method as formalised as QCA. This statement isn’t necessary though, so we have removed it.

There is also one discretionary revision (no. 11) that the authors may want to explore;

11. You write ‘This ‘dialogue’ between truth table and study reports is something which is contrary to established systematic review methods, since it explicitly encourages post-hoc reviewer interpretation’ (p18). I wonder if this issue is more related to a difference between what the synthesis actually involves, for example in meta-ethnography interpretation occurs during synthesis, whereas in meta-analysis interpretation occurs before and after synthesis (see Anderson LM, Oliver SR, Michie S, Rehfuess E, Noyes J, Shemilt I. Investigating complexity in systematic reviews of interventions by using a spectrum of methods. J Clin Epidemiol 2013; 66:1223-1229). In QCA, interpretation seems to occur during synthesis, during ‘the dialogue between truth table and study reports’. As such I’m not sure the issue is accurately compared to post-hoc statistics. It might be that this discussion would go better in the section further on, discussing types of knowledge.

Hopefully the following amendment clarifies the point we were attempting to make:

“This ‘dialogue’ is something which is contrary to established systematic review methods, since it explicitly encourages post-hoc exploration of study differences which is driven by knowledge of actual study findings. This reviewer interpretation may be more open to bias and be less replicable than a
typical sub-group analysis which aims not to deviate from a priori data extraction categories and sub-group divisions. (Though see below for a discussion of the type of knowledge generated.)"
Reviewer: Wendy Olsen

Minor Essential Revisions

Pg 3 middle is very good, but the label on Figure 1 is inadequate, and I suggest that you add some textual labels on each of the four parts to indicate what the text says they represent.

*We’ve changed it into a 2 x 2 table; we hope this makes it clearer.*

On pg 6, put the aims that are shown here into pg 2 (or somewhere, 1-3) as well, so that pg 2 presages what is to come. Readers need these structuring guides please.

*We’ve modified the sentence on page 3 where we introduce QCA to say: “We therefore examine an analytical technique, ‘Qualitative Comparative Analysis’ (QCA), which has been designed to overcome some of the limitations outlined above. Through a worked example, we demonstrate its application to systematic reviews and examine its utility when synthesising the results of complex interventions.”*

Page 7 needs to include mention of the key innovative findings. Don’t postpone them to the middle/end because by doing so, your bore your reader. We can be led to know what were a priori claims vs. what were new findings that got merged together during your analysis.

*We have amended the text on page 7 (old numbering) to read: “For the purposes of this illustration, however, we will discuss the application of QCA within the three conditions itemised above, since these map directly to our overarching conceptual framework of community engagement. As our analysis will demonstrate, this initial conceptualisation did not enable us to explain observed differences between the studies, and two additional conditions which have greater explanatory value were developed through the synthesis using QCA.”*

Pg 7 be more original. Modify the steps shown in numbered bullets as necessary to suit the systematic review function, please. Don’t tease us with possible alterations. If no alterations, then don’t mention the possibility of needing to amend the protocol.

*We have clarified the opening text to explain why alterations may be needed, but we do not yet have the data to decide one way or the other: “We will follow the six stages in the following example and, because this is an early worked example of QCA in synthesis, we will retain stage 5: consideration of ‘logical remainders’. However, this stage was not necessary in our example, and further methodological work will be required to ascertain whether this stage can be omitted from most syntheses using QCA, or whether it has a useful role to play in certain situations.”*

Pg 8 mentioning odds ratios assumes random samples were taken. Please make this explicit. The Odds Ratio (OR) section on page 8 is not preceded by any discussion of random vs. convenience samples. Please expand on pg 7 or 8 to simply distinguish studies with random vs. non random samples of cases. What you do presently say about sampling advice for QCA is correct but assumes purposive sampling, as seen in grounded theory where it is known as theoretical sampling.
We have: 1) modified the text on sampling to include ‘theoretical’: “In the context of a systematic review however, such purposive (or ‘theoretical’) sampling can be difficult to achieve, since the set of studies available may not contain both positive and negative findings relating to the same outcome”; and 2) We have added the following paragraph: “As mentioned above, transposing ‘purposive’ sampling techniques from primary research can be difficult in the context of a systematic review, because we cannot necessarily identify positive and negative cases; we can only use the studies which have evaluated the interventions and outcomes of concern. The sample of studies that we have, therefore, is more akin to an unbiased or ‘population’ sample – in that we have all the studies (that we can find) which evaluate a given intervention; and if we have little heterogeneity between results, there is little the QCA can do to help us identify sufficient conditions for success.”

And amended the opening text of the subsequent paragraph: “The outcome in this dataset is an indicator of the effectiveness of the intervention, which were all evaluated in controlled trials (both randomised and non-randomised). For the non-randomised studies, we have no reason to believe that the participants in one condition or another were more likely to breastfeed before the start of the intervention”. The rows are in descending order of consistency, which is the metric used in QCA to express how far the pattern of all the cases is consistent with sufficiency.”

Pg 12 I think you don’t mean ‘uniformity’. Here the description of QCA has gone wrong. However the next sentence is fine, and in general the description of what consistency measure represents is agreeable. So modify the sentence that has uniformity in it. Instead the consistency represents how far the pattern of all the cases (both X and Not X, and Y and Not Y, for X being the intersection of all the conditions in that configuration) is consistent with sufficiency.

Thanks! We’ve amended the text to read: “The rows are in descending order of consistency, which is the metric used in QCA to express how far the pattern of all the cases is consistent with sufficiency.”

You have erred in not giving a short paragraph on necessity. Could you add this before or after the main consistency discussion? Thank you.

We have added the following (in a systematic review we think that the identification of necessary conditions is epistemologically problematic): “We mentioned earlier that it is easier to identify sufficient conditions for an outcome than it is to find necessary conditions, because logically there may be (many) other ways of arriving at a given outcome, even if those methods are not present in any of the studies in the analysis. Bearing in mind the above mentioned problems relating to sampling – that we can do no more in a systematic review than find the studies that have already been conducted and cannot collect additional data to fill in gaps, as we might in primary research – we think it unlikely that reviewers will want to identify necessary conditions for complex interventions to generate an outcome. This is because our conceptualisation of complexity requires that we view each intervention and context as potentially unique; and therefore that a condition that may be necessary in all the studies we can see may not be necessary in all possible situations.”

1 Permitted additional discretionary change:

The interpretation of the subgroup issue is very interesting. You may want to give an example. The reader is not sure what you mean by heterogeneity of the sample, i.e. these sub groups, unless you
give 2-3 examples – one from epidemiology, one from policy studies, one from sociology of policy perhaps.

We think this is a fascinating area to develop too, but worry that we’re in danger of opening up an issue that could fill another paper. We have expanded the discussion on sub-group analysis as follows: “It is worth noting that analysing complex interventions in heterogeneous datasets usually requires sub-group analysis using statistical methods. Standard guidance on this, and views on the type of knowledge generated, varies between the very cautious “subgroups kill people”, and the pragmatic “...and lack of subgroup analysis kills people” (i.e. subgroups may differ from one another due to chance, and so be misleading; but subgroup analysis may be the only way to answer some research questions, such as the impact of social class).[23] Both may be correct in different circumstances, but it is impossible to know which pertains in any given review. They come from quite different viewpoints in the philosophy of science, with formal subgroup analysis based on deductive reasoning, and the expected warrants for making causal claims; whereas QCA can best be thought of an ‘abductive’ approach, which aims to provide an “inference to the best explanation” based on the available evidence.[24] Despite the challenges inherent in making causal inferences, decision-makers do require the types of knowledge that, at times, can only be gained from sub-group analyses; thus rejecting all such analyses may risk missing important knowledge. A QCA analysis too may suffer from the same biases as a statistical sub-group analysis, since few intervention replications are available and individual study results may be idiosyncratic and atypical. However, where useful knowledge cannot be gained through statistical synthesis, the QCA approach offers a much more formal, powerful and considered way of unpicking a complex evidence base than a simple list of individual study findings; and while different, the knowledge claims made by this type of analysis might be considered as being similar to those from a standard sub-group analysis in terms of potential bias, and potentially higher in terms of their utility and because they aim to provide explanations based on all available evidence, rather than only part of it (as is the case in a traditional sub-group analysis).”