Authors’ response to reviews

**Title:** The challenge of complexity in evaluating health policies and programs: The case of Women’s Participatory Groups to improve antenatal outcomes

**Authors:**

Sara Van Belle (svanbelle@itg.be)

Susan Rifkin (sbrifkin@gmail.com)

Bruno Marchal (BMarchal@itg.be)

**Version:** 1  **Date:** 28 Feb 2017

**Author’s response to reviews:**

BMC Health Services Research - BHSR-D-16-01235

Revision, 21 March 2016

Authors’ response to the comments of the peer reviewers

We would like to thank the peer reviewers for their relevant remarks.

Below, we explain how we took into account the comments of Reviewer 2. We also used this revision to correct a few typos and to formulate a few sentences more clearly.

Response to Mara Airoldi (Reviewer 2)

As pointed out in the abstract, the authors believe that the choice of research design should fit the nature of the question rather than being driven by methodological preferences. It would be helpful if the authors could relate their position to existing ideas on how to do this, e.g. Lucy Gilson’s methodological reader (http://www.who.int/alliance-hpsr/alliancehpsr_reader.pdf).

We completely agree with the position taken by Gilson et al. and make that point in the text on P2, first para, to which we added the sentence:

“This is in line with the current guidance for health policy and systems research [9].”

It would be helpful if the authors could say explicitly if they see any role of RCTs in assessing whether PLA produces any change in antenatal outcomes and hence if it could complement a
realist evaluation. If so, how? If not, is this an issue with PAL or with studying the impact of any complex "intervention"?

The distinction between mechanism-oriented, case-based approaches and RCTs (p5 line 18-30) brings the core argument of the paper to life. Again, it would be helpful if the authors could say if they think the two approaches can in any way complement each other in shedding light on what mechanism are at play, when and the extent to which they make a difference on outcomes of interest.

In a first response to the above comments, we better described how quantitative and qualitative data collection and analysis is used in realist evaluation. More specifically, in the example of how a RE could be set up, we added to following line (p. 7):

“A useful analytical strategy would be to start with assessing the outputs, the intermediate outcomes and the distal outcomes on the basis of a before and after study design.”

We also added a paragraph to the Discussion, because since the initial drafting of this manuscript, there has been some debate on realist RCTs

“In the meanwhile, the debate on how quasi-experimental designs can be combined (or not) with RE has started [36-39]. At the core of the debate is the issue whether the philosophical basis underpinning the (quasi-)experimental designs can accommodate a mechanism-oriented approach such as realist evaluation. More practically, the question is whether a RCT can identify mechanisms. From our perspective, the RCT design does not allow the researcher to identify the dynamic interplay between intervention, actors, contexts, mechanisms and outcomes, and thus, the term ‘realist RCT’ is a misnomer. That does not mean that RCTs have no place in research and evaluation of complex interventions. It means that the RCT should be used for what it is meant: to assess the effectiveness of interventions.”

I would also find it helpful to understand more how a mechanism-oriented approach to causal explanation can produce certainty (“the mechanism will always produce the outcome of interest”, line 22 p5) rather than a strong belief that the mechanism would produce the outcome in a similar context. Perhaps in the example of how realist evaluation would be used in the specific example of PAL, it may be helpful to read how the purposive sampling of cases and respondent would generate that certainty.

From a realist perspective, claims regarding the certainty of results will be formulated very prudently: it is through accumulation of insights from studies in different contexts that a better and more fine-grained understanding emerges. We therefore agree that the current formulation is too strong and revised the line on P. 5 as follows:

“– if it actually operates in a specific context, the mechanism will always produce the outcome of interest.”
We added a few lines on the role of purposive sampling on P7, 4th Para in the text explaining the example:

“In realist research, the cases are selected purposively: they need to enable testing of the programme theory. That means that cases are selected on the basis of variation in context, outcome or intervention.”

Finally, we also added some extra text to the end of the 5th Para related to how the programme theory is refined through a series of studies.

“The comparison of findings from the different sites would then lead to refining the initial programme theory. Ideally, this refined PT would be tested in another round of case studies. Such series of contrasting cases leads to specifying the programme theory in a gradual process, extending its explaining power.”