PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Protocol for an economic evaluation of the WHO STOPS childhood obesity stepped wedge cluster randomised controlled trial
AUTHORS	Sweeney, Rohan; Moodie, Marj; Nguyen, Phuong; Fraser, Penny; Bolton, Kristy; Brown, Andrew; Marks, Jennifer; Crooks, Nic; Strugnell, Claudia; Bell, Colin; Millar, Lynne; Orellana, Liliana; Allender, Steven

VERSION 1 – REVIEW

REVIEWER	Alastair Canaway University of Warwick, UK
REVIEW RETURNED	09-Dec-2017

GENERAL COMMENTS	This is a very interesting study, but also hugely challenging from an economic evaluation perspective. It is reassuring to see that the authors appreciate just how difficult it will be and have provided pragmatic solutions where necessary. Given the multi-faceted nature of the intervention it is impossible to lay out all the nuances that will be required within the protocol. Many issues will only arise during the study so at this stage it is impossible to address how they will deal with them. Of the 'known' issues I feel the authors have acknowledged the most significant and generally set out the means of addressing them appropriately. Below I provide a handful of minor revisions/comments. - Check the formatting of the 'Wilkie 2016' reference (p5. L46). - I'm slightly concerned about only costing components 2 and 3 for only two communities in each arm. It sounds like component three in particular could be particularly heterogenous across the different sites, so this could be an issue. However, given limited resources I understand you may have to make pragmatic decisions. It's worth keeping mind how you may deal with this should it happen. - P.13 (section 'Quality of Life'). I think you need to add more detail in relation to how you will derive your utility score. When you say you will develop an algorithm, could you give more detail? E.g. are you planning to develop and use a mapping algorithm, or are you going to do a preference elicitation study? I think the former are particularly risky (there are a lot of poor mapping algorithms in existence). - Add a sentence on how consent will be obtained. - How will the QoL data be collected? Interviewer administered, or self complete? Same for health-related behaviour. - I don't think you've mentioned what age children you are targeting, this obviously has implications for the measures that you use, e.g. which version of the PedsQL etc.

REVIEWER	Z. Zhan / GH de Bock UMCG RUG the Netherlands
REVIEW RETURNED	02-Jan-2018

GENERAL COMMENTS	This study protocol provides a detailed and well-written health economic analysis plan for a complex community-based trial named WHO STOPS, which aims to reduce childhood BMI-z score and obesity prevalence by engaging the whole communities at multiple levels. The study design of the trial is a stepped wedge cluster randomized trial with two patterns consisting of pure intervention (with baseline measurement) and one treatment switch after step 1 of the length of 2 years. 10 communities are randomized into the two patterns taking into account the possible imbalances in community size. The authors of this manuscript provided detailed descriptions regarding the complexities and complications of health economic evaluation for the trail. Indeed, due to the fact that interventions are primarily community-oriented or community-dependent and associated with a large number of decisions and datum, it is rather difficult to quantify and conduct a cost-effectiveness analysis (CEA). Nevertheless, I have two major concerns which are listed below. 1. Estimation of the intervention effect and CEA Though the statistical method with regarding how to estimate the intervention effect has been described previously (Allender et al, 2016), it is still unclear how the below-listed issues will be handled, among which some are also related to the CEA and costing calculations: a. Treatment-period interactions: First of all, as stated in the current study protocol, "current practice itself is dynamic as communities introduce new and phase out old local strategies. Implementation of the intervention in control communities will be completed two years after implementation in intervention (Step 1) communities (2019)". This means at Step1, neither the control community nor the intervention community will be under "complete" control and intervention, respectively. However, it seems that a CEA will be conducted at the beginning of Step2 relying solely on the information provided at Step 1. This approach raises concerns about the validity and biases of the analysis. Furthermore, since there are no communities under the control at Step 2, this prevents the researchers to investigate the treatment-period interactions at Step 2. b. Community-dependent interventions: Especially for component 3 of the intervention, the implemented interventions will be highly dependent on the communities, therefore the effect of the intervention will also be community-dependent. Since both intervention and control communities have a size of 5, this creates unbalances in the two arms and will very likely lead to an invalid estimation of the intervention effect. It is not clear how the authors plan to tackle this issue. Besides, according to the plan, the costing calculation will be based only on two selected communities from each arm which will results in biased estimates since the costing will be highly dependent on the two selected communities from the intervention arm even though considerations have been given about the comparability of selected intervention and control communities. c. It was mentioned in the original study protocol (Allender et al, 2016) about 13 external control communities. However, this was not mentioned in the present study protocol at all. It would be nice to know if including the external control communities in the analysis would simplify some of the questions raised above.

	2. Evaluation of future benefits Since both intervention communities and control communities will be under the intervention at Step 2 experiencing different lengths of policy changes with different community-based interventions, it is not clear how the authors plan to calculate the BMI and PA changes attributable to the intervention. Furthermore, the multi-state life-table method of Brown seems to assume a proportional hazard model, that is, the effect of the intervention (changes in BMI for instance) would have a constant effect on the disease-specific mortality proportional to the reference population. Such assumption is merely for the conveniences of statistical modeling, but are usually less true in practice since the change in childhood obesity condition could have different impacts on the health-related outcomes at different phases of life and also different between acute and chronical diseases. A more detailed elaboration is needed. Please note that due to the nature of the stepped wedge design, it is uncertain what kind of policies will be maintained at communities after the end of the trial and the impact of the intervention is generally less tractable compared to a parallel group design, especially when the double-blind is not possible. Minor suggestions: The wording "intervention communities" and "control communities" are confusing to people who are less familiar with the concept of stepped wedge design.

VERSION 1 – AUTHOR RESPONSE

Reviewers 'comments and authors' responses

Editor's changes

Editors/reviewers comments	Authors' response	Changes made
1. Please revise the title to make it clear this is a protocol e.g. "Protocol for an economic evaluation of the WHO STOPS childhood obesity stepped wedge cluster randomised controlled trial."	We have updated the title as suggested.	p.1. Title has been revised to: "Protocol for an economic evaluation of the WHO STOPS childhood obesity stepped wedge cluster randomised controlled trial"
2. Please add your dissemination plans to the 'ethics and dissemination' section of the abstract as per journal requirements for study protocols	We have dissemination plans as requested.	p.3. The following text has been added in the 'ethics and dissemination' section of the abstract. "Trial findings (including economic evaluation) will be published in peer reviewed journals and presented at international conferences. Collected data and analyses will be made available in accordance with journal policies and study ethics approvals. Results will be presented to relevant government authorities with an

		interest in cost effectiveness of these types of interventions."
3. Please revise the 'Strengths and Limitations' section 4. It shouldn't be a summary of the study (see the first two bullet points) and each point should relate to the methods/ design of the economic evaluation reported in this protocol	Thank you for providing this additional clarity. We have endeavoured to revise the bullet points and believe they now clearly relate to the study methods of the economic evaluation set out in this protocol.	p.4. The first two bullet points have been deleted. Three new bullets have been added and two significantly revised. The new bullets read: "•The protocol for this novel and complex intervention is guided by the CHEERS guidelines for economic evaluation and draws lessons from literature on the economic evaluation of complex public health interventions. •WHO STOPS will be assessed against other implementation considerations (strength of evidence, equity, acceptability to stakeholders, sustainability, feasibility of implementation, and potential side effects). •This protocol provides the core elements for the economic evaluation of this adaptive multifaceted whole of systems approach. Further decisions may be required that have not yet been considered. " The revised bullets now read: "•Pragmatic solutions are discussed for the core challenges this complex intervention poses for economic evaluation (e.g. defining intervention boundaries; measurement and attribution of costs to WHO STOPS). •This protocol provides the core elements for the economic evaluation of this adaptive

		multifaceted whole of systems approach. Further decisions may be required that have not yet been considered."
4. Please reformat the references according to the journal's guidelines.	We have amended the references according to the journal's formatting guidelines. I've not included these format changes in the tracked changes. I hope that's ok.	Changes throughout from "author-date" in text citations to reference number style [1]. Reference list formatting also amended as per BMJ style.

Reviewer 1

Editors/reviewers comments	**Authors' response**	**Changes made**
1. Given the multi-faceted nature of the intervention it is impossible to lay out all the nuances that will be required within the protocol. Many issues will only arise during the study so at this stage it is impossible to address how they will deal with them. Of the 'known' issues I feel the authors have acknowledged the most significant and generally set out the means of addressing them appropriately.	Thank you for the feedback. As you've suggested, we do anticipate more challenges to arise over time as stated in the final sentence of the manuscript (p.19) - "This does not preclude additional hurdles arising during the course of the project forcing additions to or deviations from this plan…". In response to your comment on this we've also decided to reiterate this issue as a limitation in the 'Strengths and Limitations of this Study" section.	P.4. The final bullet point now reads: "•This protocol provides the core elements for the economic evaluation of this adaptive multifaceted whole of systems approach. Further decisions may be required that have not yet been considered."
2. Check the formatting of the 'Wilkie 2016' reference (p5. L46).	Thank you for spotting. The formatting has been amended.	Changes throughout from "author-date" in text citations to reference number style [1]. Reference list formatting also amended as per BMJ style.
3. I'm slightly concerned about only costing components 2	Thanks for this comment.	p. 13, 1st paragraph in section 4.3,

	and 3 for only two communities in each arm. It sounds like component three in particular could be particularly heterogenous across the different sites, so this could be an issue. However, given limited resources I understand you may have to make pragmatic decisions. It's worth keeping mind how you may deal with this should it happen.	Evaluation resource limitations have required us to make this decision. We agree Component Three heterogeneity across communities is very possible. Whilst we think comprehensive costing of 2/5 intervention communities will likely reveal signs of heterogeneity, it is possible the presence of such heterogeneity may remain unmasked. All the other implementing communities will be provided with the Action Register template we've designed to capture the Component Three community actions and associated resource use and encouraged to use for their own monitoring and evaluation purposes. Prior to submission for publication and broader dissemination, the findings of our selected "comprehensively costed" communities will be discussed with backbone organisations from all communities. The "uncosted" community Backbone organisations will be asked to reflect on the extent to which presented findings appear broadly consisted with their own communities. Where there is a reporting of significant heterogeneity, this will be raised in the published economic evaluation results. The following text has been added to this manuscript to reflect this. To address your comments we have added new text as set out in the next column.	new text added. "A community action register, which is designed for tracking community actions and associated resource use, will be provided to each community's backbone organisation. This will facilitate data collection for evaluation purposes as well as aid each community's own monitoring and evaluation." p.15, new paragraph added. "Given the intrinsic variability of component three of the intervention it is possible that comprehensive costing of only two of the five communities might not provide an accurate representation of the large heterogeneity of actions and costs involved in this component. As a way to investigate this variability all communities will be encouraged to monitor community actions and resource contributions using the community action register. Prior to finishing the economic evaluation, results from the costing of the two selected intervention communities will be discussed with backbone organisations from all intervention communities. Those backbone organisations from communities not comprehensively costed will be asked to consider those results alongside the actions registered in their own registers, and reflect on the extent to which findings appear consistent with their own communities. If perceived heterogeneity is raised, this will be acknowledged as a limitation in the published economic evaluation results."
4.	P.13 (section 'Quality of Life'). I think you need to add more detail in relation to	Thanks for the feedback. We will use a mapping	pp.12 in the Quality of Life section – text has been revised. The

how you will derive your utility score. When you say you will develop an algorithm, could you give more detail? E.g. are you planning to develop and use a mapping algorithm, or are you going to do a preference elicitation study? I think the former are particularly risky (there are a lot of poor mapping algorithms in existence).	algorithm, which is currently being designed (by us), employing a large dataset of comparable children (1800 Australian children aged between 10 and 12 years), following the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) best practice guidelines (Wailoo et al 2106). Whilst it is true that some mapping algorithms perform poorly – we are fairly confident that a high quality mapping algorithm will be identified given: (i) the sample being used to develop the mapping algorithm come from Australian adolescents, (i.e. the same population as WHO STOPS); and (ii) the optimal mapping algorithm will be selected based on a series of econometric techniques (including Order OLS, CLAD, robust MM estimator, GLM, FMM and MFP). The mapping algorithm will be peer reviewed prior to employment in the WHO STOPS study, to further interrogate methodological rigour. For the WHO STOPS evaluation, an existing relevant algorithm published by Mpundu-Kaambwa et al. (Pharmacoeconomics 2017) will be used to conduct sensitivity analyses to test the robustness of our conclusions. All that said, we take your point that a "good algorithm" may not be identified regardless of rigour. Should it be deemed (for example, by the external peer reviewers of the Mapping	following underlined text has been added and struckthrough text deleted. "Given PedsQL is a non-preference-based HRQoL instrument, an algorithm will be developed ~~using best practice methods,~~ to enable conversion of PedsQL overall scores of study participants to the preference-based Children's Health Utility 9 Dimension (CHU-9D) index (Stevens 2012~~; Wailoo et al. 2017~~). Specifically, a dataset of around 1800 Australian children aged between 10 and 12 years will be employed. Following best practice methods, the optimal mapping algorithm will be chosen based on a series of econometric techniques using a number of goodness-of-fit measures (Wailoo et al. 2017)."

	Algorithm paper) that a good predictive model has not been identified, then we will default to an analysis whereby we would estimate the incremental cost per unit change in PedsQL scores (rather than per QALY gained). Text has been added to make the approach being taken clearer. The following underlined text has been added to reflect this in p.13 Quality of Life section.	
5. Add a sentence on how consent will be obtained	A sentence has been added.	p. 12, 1st paragraph in the 4.2 "Health and health-related behavioural outcomes" section. First sentence (with new text underlined) now reads: "Primary (BMI change) and secondary (PA and dietary behaviours) outcome data will be collected from <u>children aged between about 8 and 12 years at</u> participating primary schools <u>using opt-out (passive) consent</u> as described in Crooks et al. (2016) …."
6. How will the QoL data be collected? Interviewer administered, or self complete? Same for health-related behaviour.	This data will be self-completed, but students will be given guiding prompts on how to answer and be able to ask the research team members for help if needed. An additional sentence has been added to clarify this.	p.12, new sentence added at end of 1st paragraph in the 4.2 "Health and health-related behavioural outcomes" section. "PA and dietary behaviour questions will be self-completed, with students given structured prompts on how to answer the questions or to clarify terminology (as is the case for health-related quality of life (HRQoL) questionnaire – discussed below)."
7. I don't think you've mentioned what age children you are targeting, this obviously has implications for the measures that you use, e.g. which version of the PedsQL etc	Thank you for pointing this out. This information has been added.	See new text added in response to your query no.5.

Reviewer 2

Editors/reviewers comments	Authors' response	Changes made
1. Estimation of the intervention effect and CEA Though the statistical method with regarding how to estimate the intervention effect has been described previously (Allender et al, 2016), it is still unclear how the below-listed issues will be handled, among which some are also related to the CEA and costing calculations: a. Treatment-period interactions: First of all, as stated in the current study protocol, "current practice itself is dynamic as communities introduce new and phase out old local strategies. Implementation of the intervention in control communities will be completed two years after implementation in intervention (Step 1) communities (2019)". This means at Step1, neither the control community nor the intervention community will be under "complete" control and intervention, respectively. However, it seems that a CEA will be conducted at the beginning of Step2 relying solely on the information provided at Step 1. This approach raises concerns about the validity and biases of the analysis. Furthermore, since there are no communities under the control at Step 2, this prevents the researchers to investigate the treatment-period interactions at Step 2	Thank you to the reviewers for this feedback. We agree that this is a study limitation, given the period for which we have a "complete" (current practice) control may not be sufficient to detect a meaningful treatment effect. The duration to exposure issue you raise is also important. While we intend some level of analysis to this effect (p.11- "Results will also be analysed after four years (two years post Step 2 implementation (2021) to identify the evolution and sustainability of community responses (including resource use) and any treatment effect"), the comparison then of 4 years exposure (Step 1 communities) versus two years exposure (Step 2 communities) risks underestimating any treatment effect that might be identified against the preferred current practice control. This is a major challenge for conducting this economic evaluation in real community settings. A randomized trial required the enrolment of communities willing to adopt new strategies to tackle childhood obesity through a system approach. A parallel randomized trial was deemed not feasible and the two-step design was agreed upon in collaboration between the research team and community leaders on the basis that no participating community was	P.16-17 in section 4.5 Uncertainty and Scenario Analyses the following new paragraph has been added to the section. "It is important to note some of the challenges being faced that may have implications for the estimated treatment effect. The intervention aims to make system level changes, some of which will take longer than the trial period to occur. Further, due to the variability we have observed in the time it takes communities to reach the Component 3 stage (community actions), it is likely there will be variable effects measured across communities after two years. The four year analyses to be undertaken will be comparing a maximum of four years of exposure (Step 1 communities) versus a maximum two years of exposure (Step 2 communities) rather than comparing against a current practice control. However, BMI change data is being collected from a small number of schools in 13 communities external to the WHO STOPS study (Strugnell et al., 2016). The methods of data collection and general characteristics of the surveyed schools differ from the current study. While these external communities do not constitute the ideal control group, several of the external communities are comparable (in terms of population and geographic proximity to major urban centres). The external communities will provide an indication of BMI trends in non WHO STOPS settings. This will facilitate the estimation of plausible variations in treatment effects at 2 and 4 years for use in sensitivity

	willing to forego potentially useful “treatment” for four years (i.e. spend two-steps as a “complete” control). With this understanding, the step wedge design was proposed which, even under the limitations imposed by having only two steps, has been subjected to and approved by two important peer review processes (highly competitive National Health Medical Research Council grant funding, and the Study protocol paper). Table 1 presents a slightly simplified version of the actual implementation roll-out, for the sake of brevity for the reader. One of the communities was especially enthusiastic and ready to go and completed intervention in late 2015. The outcome pre (2015) and post GMB3 (2017) data from this community, has been analysed and compared against the Step 2 (Control communities). These have shown some promising results that give some confidence that the observation of a treatment effect after two years is plausible. It is not appropriate to make these results publicly available as yet, but they can be made available to reviewers upon request). We take your point though and these limitations have been made clearer in the manuscript with the addition	analyses.”

	of text.	
2. Community-dependent interventions: Especially for component 3 of the intervention, the implemented interventions will be highly dependent on the communities, therefore the effect of the intervention will also be community-dependent. Since both intervention and control communities have a size of 5, this creates unbalances in the two arms and will very likely lead to an invalid estimation of the intervention effect. It is not clear how the authors plan to tackle this issue.	We agree with the reviewer and appreciate your point. As described in the protocol paper (Allender et al. 2016) we will do an intention to treat analysis which will measure the effect of offering the strategy to communities with not all of them uptaking at the same pace or intensity. We acknowledge the limitations associated with measuring the effect of this type of interventions, however these are the real world conditions in which whole of system obesity interventions are developed. There is a compromise/tension between scientific rigor and real word implementation of interventions that involve communities and the numerous stakeholders that reside within them. We have now made it clear in the manuscript that an intention-to-treat analyses will be undertaken and the limitations you raise here will be acknowledged in the published results.	p.10, Section 4.1. Midway through 1st paragraph. The following sentence has been added: "Analyses of intention-to-treat and as-per-protocol (i.e. reaching Component 3) treatment effects will be undertaken."
3. Besides, according to the plan, the costing calculation will be based only on two selected communities from each arm which will result in biased estimates since the costing will be highly dependent on the two selected communities from the intervention arm even though considerations have been given about the comparability of selected intervention and control communities.	We thank the Reviewers for this comment and refer your attention to our response to Comment 3 from Reviewer 1 who shared your concerns.	Please refer to response to Comment 3 from Reviewer 1 for amendments to the manuscript text.
4. It was mentioned in the original study protocol (Allender et al, 2016) about 13 external control	We agree that employing these external communities as a contemporaneous	p.17 Uncertainty and scenario analyses section.

communities. However, this was not mentioned in the present study protocol at all. It would be nice to know if including the external control communities in the analysis would simplify some of the questions raised above.	control group would essentially add that extra time period for analysis. Unfortunately, the data from these 13 external communities are collected using a different methodology with a random sample of three schools from each local government area (which may include multiple "communities" as defined in the WHO STOPS trial) invited to participate in the data collection. Whereas the approach in the 10 WHO STOPS communities has successfully been set up as a monitoring system resulting in around 80% of schools in those communities measured, the 13 external communities surveyed only 30% of schools in the Local Government Area. This results in significantly different response rates between the 13 external and the WHO STOPS communities, with some differences in surveyed school characteristics resulting. For the aforementioned reasons the 13 external communities do not constitute an ideal control group for measuring treatment effect. However, given the limitations discussed above, data from the 13 external communities could be used to establish trends in BMI change in order to establish plausible variations in treatment effects at 2 and 4 years for use in sensitivity analyses. The following text has been added to the	"However, BMI change data is being collected from a small number of schools in 13 communities external to the WHO STOPS study.[51] The methods of data collection and general characteristics of the surveyed schools differ from the current study. While these external communities do not constitute the ideal control group, several of the external communities are comparable (in terms of population and geographic proximity to major urban centres). The external communities will provide an indication of BMI trends in non WHO STOPS settings. This will facilitate the estimation of plausible variations in treatment effects at 2 and 4 years for use in sensitivity analyses."

	manuscript to indicate this.	
5. Evaluation of future benefits. Since both intervention communities and control communities will be under the intervention at Step 2 experiencing different lengths of policy changes with different community-based interventions, it is not clear how the authors plan to calculate the BMI and PA changes attributable to the intervention.	BMI and PA change after 2 years will be estimated under an intention to treat approach with a linear mixed model for outcome (BMI or PA). The model will include intervention, time (2015, 2017, 2019) and interaction intervention x time as fixed effects and community and school as random effects. For the sake of brevity, we have not presented this detail in the manuscript rather opting to refer readers to the trial protocol itself – "…and the intervention effect assessed as described in Allender et al. (2016)" (p.13). We agree that due to variability in the commencement of component 3 of the intervention in the different step 1 communities the estimated treatment effect may be small, but as discussed above, communities were unwilling to wait for 4 years to receive the intervention. As mentioned earlier in response to your Comment 1, the outcome pre (2015) and post GMB3 (2017) data from the early implementing community shows some cause for promise of detecting a statistically significant effect at two years of exposure. We agree the study design does not readily enable the estimation of treatment	Relevant revised text already in response to your Comments 1, 2 & 4.

	effect from the longer 4 years of exposure for the reasons you've raised. As discussed in response to your previous comment, we will employ the data from the external communities in sensitivity analyses and scenario analyses to establish plausible estimations of treatment effect, with discussion of limitation clearly noted.	
6. Furthermore, the multi-state life-table method of Brown seems to assume a proportional hazard model, that is, the effect of the intervention (changes in BMI for instance) would have a constant effect on the disease-specific mortality proportional to the reference population. Such assumption is merely for the conveniences of statistical modeling, but are usually less true in practice since the change in childhood obesity condition could have different impacts on the health-related outcomes at different phases of life and also different between acute and chronical diseases. A more detailed elaboration is needed. Please note that due to the nature of the stepped wedge design, it is uncertain what kind of policies will be maintained at communities after the end of the trial and the impact of the intervention is generally less tractable compared to a parallel group design, especially when the double-blind is not possible	Thank you. We agree there is significant uncertainty around the link between any observed treatment effect during the trial period and the BMI and PA behaviours of individuals later in life. Any treatment effect observed at 2 years will form the base case for evaluating future benefits with a range of plausible assumptions of effect decay (e.g. 5% per annum) over time. These may range from 100% maintenance of effect through to full decay over 5 or 10 years, or a slowing down of the effect after a specified period. Given that by design, this system-based intervention aims to increase community response and system level change over time, it is also plausible that accentuation is possible. Should tracking of community actions via the action register, plus analyses of change in social network analyses and community readiness to change suggest accentuation appears plausible, sensitivity analyses may also include	p.17. Third paragraph is new. Newly text added reads: "Given uncertainty around the maintenance of community responses and treatment effects beyond the trial period, modelling of future benefits and health cost savings will test a range of plausible assumptions of decayed and maintained treatment effect. These may range from full decay over 5 or 10 years through to 100% maintenance of effect. In the event that within trial analyses of the proxy indicators of system change suggest an accentuation of treatment effect is plausible, modest accentuation of treatment effect will also be modelled."

	modelling of a modest accentuation of effect. Variation of these assumptions of effect size decay and accentuation will be used in a threshold analysis to determine the values of these (and other) important variables at which WHO STOPS reaches key thresholds of cost-effectiveness. We have added some text to Section 4.5 on Uncertainty and scenario analyses to more clearly state our plans to address this uncertainty.	
7. The wording "intervention communities" and "control communities" are confusing to people who are less familiar with the concept of stepped wedge design	Thanks for the feedback. Intervention and control are standard terms to describe two arms of a controlled trial in health and we feel that readers will be more comfortable with differentiating the two arms of our trial with the terms "Intervention Communities" and "Control communities", rather than Step 1 and Step 2 communities. We've tried to make this clear in section 2.1 Study Design where we state: "In brief, ten dispersed clusters or 'natural communities'…..were randomly assigned to receive the intervention at Step 1 (2017) – referred to as intervention communities, or Step 2 (2019) referred to as control communities"	

VERSION 2 – REVIEW

REVIEWER	Alastair Canaway University of Warwick

REVIEW RETURNED	14-Feb-2018

GENERAL COMMENTS	The authors have thoroughly addressed my previous comments and I am satisfied with their response.

REVIEWER	GH de Bock and Z Zhan Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, Eindhoven, The Netherlands
REVIEW RETURNED	16-Feb-2018

GENERAL COMMENTS	Nice work