PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A Theory of Change for the delivery of talking therapies by lay workers to survivors of humanitarian crises in low- and middle-income countries: Protocol of a systematic review
AUTHORS	Ryan, Grace; Bauer, Andreas; Bass, Judith; Eaton, Julian

VERSION 1 – REVIEW

REVIEWER	Brian Guthrie Ph D Mount Royal University Calgary, AB. Canada
REVIEW RETURNED	26-Jun-2017

GENERAL COMMENTS	This study will be an important contribution to advancing both the knowledge and the practice of developing, supporting and training non specialist mental health counselors in LMIC. In reviewing the proposal I am reminded of Harare's park bench grandmothers Located in Highfield, a poor suburb just south of Zimbabwe's capital Harare, they are trained but unqualified health workers who take turns on the park bench to hear stories (Friendship Bench project). I look forward to reading the results of your study.

REVIEWER	MariannaPurgato University of Verona, Italy I published two papers with author Judy Bass (An ecological model for refugee mental health: implications for research. Purgato M, Tol WA, Bass JK. Epidemiol Psychiatr Sci. 2017 Apr;26(2):139-141; Mental health and psychosocial support in humanitarian settings: a public mental health perspective.Tol WA, Purgato M, Bass JK, Galappatti A, Eaton W. Epidemiol Psychiatr Sci. 2015 Dec;24(6):484-94)
REVIEW RETURNED	02-Oct-2017

GENERAL COMMENTS	The paper presents a protocol for a systematic review focused on the critical issue of talking therapies delivered by lay counselors in low-and-middle-income countries. I read with interest the manuscript. However, I have the following concerns: 1. Data collection process: please, explain how different studies (qualitative, quantitative, and mixed methods) will be managed in the development of the review, and what type of outcome measures will be collected (perhaps with different instruments).

	2. Risk of bias in individual studies: you stated that “studies will not be excluded from the synthesis on the basis of the quality”. I suggest adding a sensitivity analysis (for quantitative studies) excluding low-quality studies, to check if this affects results. 3. Data synthesis paragraph: please, provide more detail on data synthesis and analysis. How data related on different common mental disorders will be considered and analyzed? How studies involving mixed populations (i.e., adults with comorbidities, or studies involving participants with different mental disorders) will be considered? Please, specify if you intend to consider some crucial variables such as the time since trauma exposure, the type of trauma, the previous/current exposure to other treatments (either psychological or pharmacological), and the format of therapy (individual versus group). 4. Related to point 3, I would suggest adding a specific paragraph dedicated to data analyses for quantitative studies assessing effectiveness, specifying subgroup and/or sensitivity analyses.

REVIEWER	Anna Chiumento University of Liverpool, UK
REVIEW RETURNED	06-Oct-2017

GENERAL COMMENTS	This is a very clearly thought through and presented protocol paper which I enjoyed reading and recommend for publication. The study addresses the important aim of developing a theory of change to identify the change process behind lay-delivered talking therapies in humanitarian settings, which in turn will facilitate identifying where current evidence is weak or incomplete. My comments are for the authors to consider making very minor ammendments to the manuscript. These are broken down by paper section for ease: RATIONALE: - In the introduction some of the language feels sensationalist (e.g. "reeling from disaster" and often overly complex (e.g. "putative variables), consider simplifying for an audience whose first language may not be English and to be more neutral. - Definitions: of "humanitarian crisis" for the review comes a little way into the paper, consider adding a brief explanation in the rationale section, you also use the term "post-disaster" which it is not clear what this refers to; equally for lay worker. - Line 26-30 the sentence talks about current research as well as the development of guidelines, could some references relating to research efforts be included here? - I would recommend adding a brief explanation of the Theory of Change for audiences who may not be familiar with this approach, and explanation of key terms. Perhaps reproducing the ToC table from the de Silva et al (2014) paper could be useful here? OBJECTIVE: - AIM: The aim statement seems to broad, specifically the mention of how implementation affects outcomes. My understanding from the rest of the paper is that the study aims to develop a common theory of change for lay-delivered evidence-based talking therapies, and suggest avenues for future research. As under the research questions, the aim is to develop a ToC of how programmes seek to achieve stated outcomes, not to explore how the implementation affects outcomes.

	I strongly suggest revising the overall aim to be more clearly linked to the activities to be undertaken in this systematic review. METHODS / ANALYSIS: - line 11: change "was" to "is" so it reads "heterogenity IS observed...." - Line 25, item 2: "How do the interventions and indicators linked to each step on this pathway vary across studies?" I find this question a bit vague and am struggling to understand how gathering information on variance will tell you much, beyond suggesting contextual differences. I may be misunderstanding, in which case i suggest re-writing this question to be clearer. INTERVENTIONS: - p.6, line 3; I suggest revising this sentence to "We include evidence-based theapeutic talking theapies....." etc to be more in line with what is discussed in this section about the types of interventions included. - p.6, line 19-21: I'm not sure I follow this sentence very well. I don't see how studies with undiagnosed CMD could claim to be aiming to treat a CMD - it seems to me to aim to treat something you need to know someone is experiencing it. Sub-threshold CMD is fine, it is just the undiagonsed aspect I am confused by. - p.7, line 22: Sub-heading "report characteristics" should be in italics. I would also be interested to know how the reviewer team anticipate reviewing any articles NOT in English (e.g. what are the language skills within the review team / intention to recruit and train additional reviewers / interntion to translate and then review papers, etc?) INFORMATION SOURCES: - Can a rationale for the 2 specific target journals be provided please? - Contact with experts: I wonder if these "experts" could be retained as an advisory group to complement input from study authors on the final ToC? It would enhance the validity of findings to have them assessed by those outside of researchers / authors of included studies. RISK OF BIAS: - p.9, line 7: reference is made to methodological quality informing narrative synthesis - can it be clarified how this will happen? DATA ANALYSIS: - It may be too late, but could the process of synthesis be revised such that a sub-section of included studies be used to develop an initial ToC, which is then revised and refined as further synthesis is undertaken? This appears to have the advantage of giving the authors something for the review to aggitate against, and reveal where and why the ToC may be being revised. This point about the iterative evolution of the ToC as a "living document" is made at the end of this section, and seems to be an appropriate approach for this study to adopt also so as not to "write out" important decisions made in the process of developing the initial ToC. PRELIMINARY SYNTHESIS: - P.9, line 44 - can a reference be provided for conducing summative content analysis?

	- p.9, line 47-48: instead of "transforming" qualitative data into quantiative data, could the authors refer to this process as "analysing" the qualitaitve data quantatively? This removes the implicit suggestion that the data is better expressed qualitatively. Can an explicit section on study strengths and limitations be included in the main text, beyond bullet points in the abstract? Also, in reference to the last comment on transferrability / generalizability, could this be placed in the context of the study which aims to develop a common ToC which would then always need to be locally refined and applied, and thus is never generalizable, but always transferrable? Can I also suggest that the authors consider the inherent limitatation behind using peer-review publications of intervention studies as the basis for identifying the theory of change behind interventions. I would suggest a short concluding section to summarise the role and importance of this systematic review to draw the paper to a close. I hope these comments assist the authors in further enhancing this protocol paper.

REVIEWER	Mukdarut Bangpan UCL Institute of Education, United Kingdom
REVIEW RETURNED	07-Oct-2017

GENERAL COMMENTS	Feedback • This is a very important topic for systematic review with the aim to develop a common framework or Theory of Change of the intervention. However, the protocol needs to be more coherent, and provide further details explaining the processes at various systematic review stages. Overall, it is not clear how the review will address the proposed research questions, particularly about implementation research. • There has been an increasing evidence base in the field of mental health and psychosocial support in humanitarian crisis. The scope review identified at least 15 reviews (Bangpan et al 2015). It would be helpful for the authors to clearly justify in the protocol how this review will fill the current gaps in research and knowledge in the field. • P.4 Line 45-47. Please unpack the statement - it is unclear how by examining similarities and differences (please explain further- similarities and differences in what aspects) will help to identify 'putative variables' that might influence outcomes at each step among this (which?) theoretical pathway of change. • P.5- primary objective: To conduct a narrative synthesis of the published and unpublished literature on lay-delivered therapies CMDs, among survivors of humanitarian crises in LMICs. However, in the review questions, it seems that it focuses more on ToC which is the secondary objective? • p.5 Line 41-54 Population: Please justify why the review will include only studies focusing on treatment for adults? Nearly 250 million Children are affected by humanitarian crisis and in fact they are the most vulnerable population groups during crises (UNICEF, 2016)

	• P.6 Line 23-24 – The authors state “General psychoeducation or psychosocial interventions with no evidence-based therapeutic component (e,g. Psychosocial First Aid)… will be excluded.” It is unclear how the team will decide which therapeutic component is evidence based? The team provides criteria for probably efficacious treatments (Table 1) arguing that the intervention should at minimum meet one of the three criteria for probably efficacious treatments. Does this mean that the team will be assessing the efficacy of each therapeutic intervention for making decisions whether to include some talking therapies? This process needs to be clearly explained how the decisions will be made. For example, if PFA evidence is mixed and inconclusive, how this will inform the decisions about inclusion of relevant PFA studies. • p.6 and p.7 about outcomes (please see also search strategy)- for interventions to be included in the review “…interventions should target one or more CMDs.” For the case that this specific information is not reported in the studies. What is the plan for applying this inclusion criterion? At the same time, on the outcome criterion on p.7 the authors state that ‘studies must report one or more client outcomes (e.g. satisfaction, symptoms or functioning) or implementation outcomes related to the treatment of CMDs”. I wonder this would exclude studies that evaluated interventions that may not report or target CMDs but assessed relevant and useful client’s and implementation outcomes. Please explain about this further. • P.7 line 4 : implementation outcomes- could you provide a more complete list/define what outcomes would be ‘implementation outcomes’ and would not? Is it penetration is the same as coverage? What do you mean by ‘fidelity’, ‘adoption’ or ‘adaption’ or they are the same? • P.7 line6-7, Could you please explain further how the outcomes will be measured qualitatively? • P.7 search strategy: It is unclear in terms of inclusion criteria and search strategy- a study will be excluded from the review if not assessed CMD outcomes? If not, CMDs should not be in the search strategy. (see also the previous comment on outcomes) • P.8 data collection process and data items- the authors refer to main ToC elements- outcomes, interventions, indicators, rationales, and assumptions, which I think they at very high level of ToC. I think it is more about presentation of the protocol- there is information about ToC but in the data items (line 46-47). Another point, the authors describe the process of extracting data from ‘a brief narrative summary’. From my understanding, the authors aim to identify a brief narrative summary to describe the pathway to changes linking these elements (outcomes, interventions, indicators etc.). Can you please specify what ‘a brief narrative summary’ would be? Is it a finding from each study, study authors’ conclusion and discussion etc.? • Please the team provide a draft version of data extraction tool in the appendix. • What is the plan of how to extract quantitative data (effect size or percentage?) and mixed methods data? • P.8 line 41-42, can you explain further the terms ‘evidence of effect’ and ‘evidence of effect modification’ • P.8 line 46-51- ‘Additional qualitative data will be extracted to inform the development of ToC.’ –Is this the same as a brief narrative summary discussed in the data collection process? (see also my comment about ToC).

	In addition, I am surprised that contexts and populations where the interventions are delivered (e.g. types of humanitarian crisis, levels of trauma/symptoms of populations at baseline), are not included/mentioned in these domains/ToC. • P.9 line 29-30- there is no discussion on 'existing' ToC. • The authors imply in many occasions that implementation, implementation research/ implementation outcomes is a focus of this systematic review (p.4 line41, line54-55, p.5 line 13-15, p5 line 28-29). It would be helpful in the background section to discuss on implementation research how it is important and relevant to this review and in the field of humanitarian crisis in general. • Overall, It is unclear how the quantitative studies would contribute to the synthesis, addressing the review questions. Please elaborate further in the method section • Risk of bias- Can you provide more details of the frameworks (which criteria of quality will be considered) and how the team will judge the quality of the studies in the review (low, medium, high risk of bias etc). This could be presented in the appendix. • Data synthesis- The authors state that- 'there is a risk that a ToC developed at an early stage of this review would be based largely on the reviewers' incomplete, a priori understanding'. However, I think it is important in terms of transparency for the review to identify the ToC known in the field (for example in mental health but not in humanitarian settings) which could guide data collection as well as data synthesis process. In my comments on data collection process and data item, the authors need to be clearer of how the team will extract what data and I think identifying/developing initial ToC based on relevant frameworks would facilitate data collection and data synthesis processes. • P.9 line 41-53- Preliminary synthesis- what the main findings from the preliminary synthesis would be, only summary statistics to describe what elements? Please provide more details. • P.9 Exploring relationships- it is unclear about this process. Are the authors planning to explore relationships between key elements of ToC? What summary statistics will be used to explore relationships? Can you please explain more about points of heterogeneity within interventions and how the team will assess potential effect modifiers? • P.10 Assessing robustness- The team will assess TOC according to study design (please provide examples of this) and the score assigned during the quality appraisal of the primary studies (please explain further about the score in the quality assessment section) • In many cases, programmes designed for populations affected by humanitarian crises involves 'talking therapies' and/or with a combination of other treatments. Please provide examples of talking therapy programmes that might be relevant to this review or other programmes that may be excluded.

VERSION 1 – AUTHOR RESPONSE

5. INTRODUCTION

• Reviewer 4 Comment 5.1: There has been an increasing evidence base in the field of mental health and psychosocial support in humanitarian crisis. The scope review identified at least 15 reviews

(Bangpan et al 2015). It would be helpful for the authors to clearly justify in the protocol how this review will fill the current gaps in research and knowledge in the field.

• Authors Reply 5.4: We have added more detail on the existing literature and justification for this review. As this has added substantially to the length of the Introduction, we have divided it into “Background”, “Rationale” and “Aims and Objectives” with sub-headings, so that there is clear sign-posting for the reader (pp. 3-4)

• Reviewer 3 Comment 5.2: In the introduction some of the language feels sensationalist (e.g. "reeling from disaster" and often overly complex (e.g. "putative variables), consider simplifying for an audience whose first language may not be English and to be more neutral.

• Reviewer 3 Comment 5.3: Definitions: of "humanitarian crisis" for the review comes a little way into the paper, consider adding a brief explanation in the rationale section, you also use the term "post-disaster" which it is not clear what this refers to; equally for lay worker.

• Reviewer 3 Comment 5.4: Line 26-30 the sentence talks about current research as well as the development of guidelines, could some references relating to research efforts be included here?

• Authors Reply 5.2-5.4: We have revised this section for tone and clarity (pp. 3-4) and moved up the definitions for “humanitarian crisis” and “talking therapy”. We have also replaced “post-disaster” with “other crises”. (pp.3). As there is some ambiguity in the literature around what constitutes a “protracted crisis”, we have added addition detail to our eligibility criteria (pp. 4). Also, we have added an example of recent research: PM+ trials in Kenya and Pakistan (pp. 3)

• Reviewer 3 Comment 5.5: The aim statement seems to broad, specifically the mention of how implementation affects outcomes. My understanding from the rest of the paper is that the study aims to develop a common theory of change for lay-delivered evidence-based talking therapies, and suggest avenues for future research. As under the research questions, the aim is to develop a ToC of how programmes seek to achieve stated outcomes, not to explore how the implementation affects outcomes. I strongly suggest revising the overall aim to be more clearly linked to the activities to be undertaken in this systematic review.

• Reviewer 4 Comment 5.6: This is a very important topic for systematic review with the aim to develop a common framework or Theory of Change of the intervention. However, the protocol needs to be more coherent, and provide further details explaining the processes at various systematic review stages. Overall, it is not clear how the review will address the proposed research questions, particularly about implementation research.

• Reviewer 4 Comment 5.7: P.5- primary objective: To conduct a narrative synthesis of the published and unpublished literature on lay-delivered therapies CMDs, among survivors of humanitarian crises in LMICs. However, in the review questions, it seems that it focuses more on ToC which is the secondary objective?

• Reviewer 4 Comment 5.8: The authors imply in many occasions that implementation, implementation research/ implementation outcomes is a focus of this systematic review (p.4 line41, line54-55, p.5 line 13-15, p5 line 28-29). It would be helpful in the background section to discuss on implementation research how it is important and relevant to this review and in the field of humanitarian crisis in general.

• Reviewer 3 Comment 5.9: Line 25, item 2: "How do the interventions and indicators linked to each step on this pathway vary across studies?" I find this question a bit vague and am struggling to

understand how gathering information on variance will tell you much, beyond suggesting contextual differences. I may be misunderstanding, in which case i suggest re-writing this question to be clearer.

• Reviewer 3 Comment 6.0: line 11: change "was" to "is" so it reads "heterogenity IS observed...."

• Authors Reply 5.5-6.0: These were extremely helpful recommendation; we have amended the aims and objectives accordingly (pp.4) and have made other changes throughout to de-emphasise the focus on implementation research that was confusing reviewers. There is no longer any mention of “heterogeneity” or “putative variables”, though we do provide more detail on the sorts of similarities and difference we expect to see and why it is important to examine these (pp.4). Notably, we have also removed the research questions, which, after revising the statement of aims and objectives, were made redundant. This has also helped to save word count.

6. METHODS/ANALYSIS

6A. Eligibility Criteria

• Reviewer 4 Comment 6A.1: 4: p.5 Line 41-54 Population: Please justify why the review will include only studies focusing on treatment for adults? Nearly 250 million Children are affected byhumanitarian crisis and in fact they are the most vulnerable population groups during crises (UNICEF, 2016)

• Authors Reply 6A.1: We have added an explanation for exclusion of several vulnerable populations in the limitations section (pp. 11).

• Reviewer 3 Comment 6A.2: p.6, line 3; I suggest revising this sentence to "We include evidence-based theapeutic talking theapies....." etc to be more in line with what is discussed in this section about the types of interventions included.

• Authors Reply 6A.2: We have revised the sentence and added examples of the types of interventions included and excluded (pp.5)

• Reviewer 4 Comment 6A.3: P.6 Line 23-24 – The authors state “General psychoeducation or psychosocial interventions with no evidence-based therapeutic component (e,g. Psychosocial First Aid)… will be excluded.” It is unclear how the team will decide which therapeutic component is evidence based? The team provides criteria for probably efficacious treatments (Table 1) arguing that the intervention should at minimum meet one of the three criteria for probably efficacious treatments. Does this mean that the team will be assessing the efficacy of each therapeutic intervention for making decisions whether to include some talking therapies? This process needs to be clearly explained how the decisions will be made. For example, if PFA evidence is mixed and inconclusive, how this will inform the decisions about inclusion of relevant PFA studies.

• Authors Reply 6A.3: Many studies cite systematic reviews and/or individual efficacy/effectiveness studies when describing the intervention; therefore, it is often possible to identify a “probably efficacious” treatment during full-text screening. We have added more detail on how the review team will investigate when this is not the case (pp. 5). It is worth noting that the criteria used by Chambless et al. do not take into consideration the actual strength of the evidence for efficacy (i.e. whether or not the evidence is mixed or inconclusive), which is why the “probably” is added. We have also reworded to make it clearer that PFA is an example of an intervention that we would exclude, not an example of an “evidence-based therapeutic component”.

• Reviewer 3 Comment 6A.4: p.6, line 19-21: I'm not sure I follow this sentence very well. I don't see how studies with undiagnosed CMD could claim to be aiming to treat a CMD - it seems to me to aim to treat something you need to know someone is experiencing it. Sub-threshold CMD is fine, it is just the undiagonsed aspect I am confused by.

• Reviewer 4 Comment 6A.5: p.6 and p.7: about outcomes (please see also search strategy)- for interventions to be included in the review "…interventions should target one or more CMDs." For the case that this specific information is not reported in the studies. What is the plan for applying this inclusion criterion? At the same time, on the outcome criterion on p.7 the authors state that 'studies must report one or more client outcomes (e.g. satisfaction, symptoms or functioning) or implementation outcomes related to the treatment of CMDs". I wonder this would exclude studies that evaluated interventions that may not report or target CMDs but assessed relevant and useful client's and implementation outcomes. Please explain about this further.

• Reviewer 4 Comment 6A.6: P.7 search strategy: It is unclear in terms of inclusion criteria and search strategy- a study will be excluded from the review if not assessed CMD outcomes? If not, CMDs should not be in the search strategy. (see also the previous comment on outcomes)

• Authors Reply 6A.4-6A.6: The intervention must be delivered for the treatment of CMDs, and this is specified in the inclusion criteria and search terms; however, we are including implementation research which might not assess CMD outcomes. We believe it is now clear from the text that studies which do not report or target CMDs will be excluded on the basis that they do not represent an intervention of interest. We have removed undiagnosed CMD but will continue to include sub-threshold cases; we have adjusted the text here accordingly. (pp. 5)

• Reviewer 4 Comment 6A.7: line 4: implementation outcomes- could you provide a more complete list/define what outcomes would be 'implementation outcomes' and would not? Is it penetration is the same as coverage? What do you mean by 'fidelity', 'adoption' or 'adaption' or they are the same?

• Authors Reply 6A.7: We have added a table (Table 2) which provides this information from Proctor et al. (2011). (pp.6).

• Reviewer 4 Comment 6A.8: P.7 line6-7, Could you please explain further how the outcomes will be measured qualitatively?

• Authors Reply 6A.8: Outcomes are not measured qualitatively, but qualitative data about outcomes will be extracted. We have changed the sentence to read: "Outcomes may be measured quantitatively or described qualitatively" (pp. 6).

• Reviewer 3 Comment 6A.9: p.7, line 22: Sub-heading "report characteristics" should be in italics. I would also be interested to know how the reviewer team anticipate reviewing any articles NOT in English (e.g. what are the language skills within the review team / intention to recruit and train additional reviewers / interntion to translate and then review papers, etc?)
• Authors Reply: We have changed the sub-heading accordingly and provided more detail r.e. non-English studies (pp. 7).

6B. Information Sources

• Reviewer 3 Comment 6B.1: Can a rationale for the 2 specific target journals be provided please?
o Authors Reply 6B.1: As hand searches are quite time-intensive, we aimed to choose one each of a journal focussed on humanitarian crises (Conflict & Health), a journal focussed on mental health systems strengthening (IJMHS) and a journal focussed on clinical interventions (World Psychiatry). If

the editors feel it is worth discussing in the manuscript, we are happy to provide this added detail; however, we are aware that the word count is quite high and protocols do not always explain the rationale behind the choice of databases, journals etc. Therefore we have chosen to omit it here.

• Reviewer 3 Comment 6B.2: Contact with experts: I wonder if these "experts" could be retained as an advisory group to complement input from study authors on the final ToC? It would enhance the validity of findings to have them assessed by those outside of researchers / authors of included studies.
o Authors Reply 6B.2: We have incorporated this recommendation under “assessing robustness” (pp. 11).
6C. Study Records and Data Items

• Reviewer 4 Comment 6C.1: P.8 data collection process and data items- the authors refer to main ToC elements- outcomes, interventions, indicators, rationales, and assumptions, which I think they at very high level of ToC. I think it is more about presentation of the protocol-there is information about ToC but in the data items (line 46-47). Another point, the authors describe the process of extracting data from ‘a brief narrative summary’. From my understanding, the authors aim to identify a brief narrative summary to describe the pathway to changes linking these elements (outcomes, interventions, indicators etc.). Can you please specify what ‘a brief narrative summary’ would be? Is it a finding from each study, study authors’ conclusion and discussion etc.?

• Reviewer 3 Comment 6C.2: I would recommend adding a brief explanation of the Theory of Change for audiences who may not be familiar with this approach, and explanation of key terms. Perhaps reproducing the ToC table from the de Silva et al (2014) paper could be useful here?

• Reviewer 4 Comment 6C.3: Please the team provide a draft version of data extraction tool in the appendix.

• Reviewer 4 Comment 6C.4: What is the plan of how to extract quantitative data (effect size or percentage?) and mixed methods data?

• Reviewer 4 Comment 6C.5: P.8 line 41-42, can you explain further the terms ‘evidence of effect’ and ‘evidence of effect modification’

• Authors Reply 6C.1-6C.5: Building off of other recommendations to revise the methods of analysis to more closely follow the process outlined by Popay et al. (i.e. starting with a draft ToC and amending iteratively), we have changed the process for extracting data, rendering most of these points moot. The data extraction tool, for example, will be based on the initial working ToC which has not yet been agreed, so is not yet available. It is no longer necessary to produce narrative summaries of each study to code inductively, in order to identify a common framework, if we are already using a working ToC based on previous research.

We have provided more information on ToC—including the recommended table—and edited for clarity r,e. the qualitative nature of this synthesis, which will treat all data (including from quantitative and mixed-methods studies) as qualitative data for content analysis (pp. 9-10). There is also no longer reference to “evidence of effect” or “evidence of effect modification”.

• Reviewer 4 Comment 6C.6: ‘Additional qualitative data will be extracted to inform the development of ToC.’ –Is this the same as a brief narrative summary discussed in the data collection process? (see also my comment about ToC). In addition, I am surprised that contexts and populations where the interventions are delivered (e.g. types of humanitarian crisis, levels of trauma/symptoms of populations at baseline), are not included/mentioned in these domains/ToC.

• Authors Reply 6C.6: In the process of conducting the content analysis, we expect to identify similarities and differences in the included texts, which will almost certainly cover the contexts and populations in which the interventions are delivered. These are not necessarily pre-defined and can emerge from the data in the process of conducting the synthesis. However, we are concerned it is not an efficient use of the tight word count to speculate about what will ultimately be included in the ToC, particularly as we have now chosen to adapt the Theory of Change proposed by Bangpan et al. which does very clearly highlight these variables already (pp. 10). We have also tried to highlight the importance of accounting for these similarities and differences in the rationale and objectives (pp. 4).

6D. Study Quality

Reviewer 2 Comment 6D.1: Risk of bias in individual studies: you stated that “studies will not be excluded from the synthesis on the basis of the quality”. I suggest adding a sensitivity analysis (for quantitative studies) excluding low-quality studies, to check if this affects results.

• Authors Reply: Based on the results of our scoping review, we do not expect to identify a sufficient number of quantitative studies suitable for meta-analysis, and therefore will not be conducting a sensitivity analysis. We have tried to state this more clearly in the manuscript (e.g. “Limitations”, pp.11).

• Reviewer 4 Comment 6D.2: Risk of bias- Can you provide more details of the frameworks (which criteria of quality will be considered) and how the team will judge the quality of the studies in the review (low, medium, high risk of bias etc). This could be presented in the appendix.

• Reviewer 3 Comment 6D.3: - p.9, line 7: reference is made to methodological quality informing narrative synthesis - can it be clarified how this will happen?

• Authors Reply 6D.2-6D.3: We have provided additional detail on quality assessment (pp.8) and on how study quality will be taken into account in the synthesis (“Assessing Robustness”, pp. 11); however, we have not produced an additional web appendix, as the tools used are all freely available online and can be accessed through the URLs in the references section.

6E. Data synthesis

• Reviewer 3 Comment 6E.1: It may be too late, but could the process of synthesis be revised such that a sub-section of included studies be used to develop an initial ToC, which is then revised and refined as further synthesis is undertaken? This appears to have the advantage of giving the authors something for the review to aggitate against, and reveal where and why the ToC may be being revised. This point about the iterative evolution of the ToC as a "living document" is made at the end of this section, and seems to be an appropriate approach for this study to adopt also so as not to "write out" important decisions made in the process of developing the initial ToC.
• Reviewer 4 Comment 6E.2: P.9 line 29-30- there is no discussion on ‘existing’ ToC.

• Reviewer 4 Comment 6E.3: Data synthesis- The authors state that- ‘there is a risk that a ToC developed at an early stage of this review would be based largely on the reviewers’ incomplete, a priori understanding’. However, I think it is important in terms of transparency for the review to identify the ToC known in the field (for example in mental health but not in humanitarian settings) which could guide data collection as well as data synthesis process. In my comments on data collection process and data item, the authors need to be clearer of how the team will extract what data and I think identifying/developing initial ToC based on relevant frameworks would facilitate data collection and data synthesis processes.

• Authors Reply 6E.1-6E.3: We have revised the synthesis to better align with the methods proposed by Popay et al. in response to the reviewers' recommendation that a working ToC be adopted at an early stage of the review, and specifically that this build on existing ToCs (pp. 9-11). The existing ToC by Bangpan et al is now used as the initial point of departure for ToC development (pp. 10).

• Reviewer 2 Comment 6E.4: Data synthesis paragraph: please, provide more detail on data synthesis and analysis. How data related on different common mental disorders will be considered and analyzed? How studies involving mixed populations (i.e., adults with comorbidities, or studies involving participants with different mental disorders) will be considered? Please, specify if you intend to consider some crucial variables such as the time since trauma exposure, the type of trauma, the previous/current exposure to other treatments (either psychological or pharmacological), and the format of therapy (individual versus group).

• Reviewer 2 Comment 6E.5: Related to point 3, I would suggest adding a specific paragraph dedicated to data analyses for quantitative studies assessing effectiveness, specifying subgroup and/or sensitivity analyses.

• Reviewer 2 Comment 6E.6: Data collection process: please, explain how different studies (qualitative, quantitative, and mixed methods) will be managed in the development of the review, and what type of outcome measures will be collected (perhaps with different instruments).

• Reviewer 4 Comment 6E.7: Overall, It is unclear how the quantitative studies would contribute to the synthesis, addressing the review questions. Please elaborate further in the method section

• Authors Reply 6E.4-6E.7: As described above, based on the results of our scoping review, we do not expect to identify a sufficient number of quantitative studies suitable for meta-analysis. We have tried to state this more clearly in the manuscript (e.g. "Limitations", pp.11). We have also included more detail about the outcomes of interest (pp. 6), data collection (pp. 8) and methods of synthesis and analysis (pp. 8-11) in this draft. We believe it is now clear from these changes that the narrative synthesis relies primarily on qualitative methods. Even quantitative outcomes of included studies may be analysed qualitatively using methods of content analysis, for example by coding the results of a study. The data items to be extracted are not all pre-defined, by design; these decisions will be made based on the Theory of Change and on the studies that are identified by the literature search.

• Reviewer 3 Comment 6E.8: P.9, line 44 - can a reference be provided for conducing summative content analysis?

• Reviewer 3 Comment 6E.9: p.9, line 47-48: instead of "transforming" qualitative data into quantiative data, could the authors refer to this process as "analysing" the qualitaitve data quantatively? This removes the implicit suggestion that the data is better expressed qualitatively.
• Authors Reply 6E.8-6E.9: We have added the citation for summative content analysis. "Transforming data onto a common rubric" is an important technique for synthesis described in the guidance by Popay et al (2006); therefore, we are hesitant to change this wording.

7. LIMITATIONS

• Reviewer 3 Comment 7.1: Can an explicit section on study strengths and limitations be included in the main text, beyond bullet points in the abstract? Also, in reference to the last comment on transferrability / generalizability, could this be placed in the context of the study which aims to develop a common ToC which would then always need to be locally refined and applied, and thus is never generalizable, but always transferrable? Can I also suggest that the authors consider the inherent

limitatation behind using peer-review publications of intervention studies as the basis for identifying the theory of change behind interventions.
• Reviewer 3 Comment 7.2: I would suggest a short concluding section to summarise the role and importance of this systematic review to draw the paper to a close.
o Authors Reply 7.1-7.2: These are very welcome suggestions and we have amended the bullet points and added a limitations section to address the reviewer's concerns. However, we could not accommodate an additional section for the conclusion, given the word-count.

VERSION 2 – REVIEW

REVIEWER	Marianna Purgato University of Verona, Italy I published two papers with author Judith Bass
REVIEW RETURNED	09-Nov-2017

GENERAL COMMENTS	The manuscript reads well and has been improved according to peer reviewers comments.

REVIEWER	Anna Chiumento University of Liverpool, UK
REVIEW RETURNED	22-Nov-2017

GENERAL COMMENTS	I would like to thank the authors for their efforts to resubmit a vastly improved manuscript that I very much enjoyed reading. The authors have put considerable effort into addressing the comments from reviewers - many thanks for your comprehensive response to each of the points raised. I would like to particularly recognise the improvements in the manuscript in terms of clearly outlining the rationale and justification for the overall study, as well as each individual step involved in this review. I also found the additional description on how the ToC will be developed very informative and easy to follow. I have no hesitation in recommending this paper for publication. I have some very minor comments on the main mansucript that the authors may wish to address: - p. 3, lines 48-50 where you describe a lay worker as "with no tertiary education or formal professional or paraprofessional certification in mental health" this implies that the non-specialist will have NONE of these attributes. Similarly on p.5 lines 23-25, the definition from Lewin et al also reinforced not having tertiary education. However, in a number of studies (e.g. Group PM+ in Swat) the "lay-workers" had batchelor degrees. Therefore, based on the above criteria I assume this study would be excluded? I wonder if this has been fully considered by the author team? (To note, the justification for this approach in PM+ is that this is a vast untapped group of appropriate lay workers in the Pakistani setting). - The last line on p.3 doesn't read very clearly, how about "Through this initiative, Porblem Management Plus has since been manualised and trialled in Kenya and Pakistan using lay counsellors (refs)"?

	- p.11, limitations section, last sentence of the last paragraph (line 43-45) doesn't make sense in the context of what has just been said, please review and revise. I very much look forward to seeing the results from this systematic review in the near future!

REVIEWER	Mukdarut Bangpan UCL, UK
REVIEW RETURNED	26-Nov-2017

GENERAL COMMENTS	Much improved manuscript. Great opportunities to understand further on the impact and of interventions delivered by lay workers in the humanitarian settings.

VERSION 2 – AUTHOR RESPONSE

-p. 3, lines 48-50: As per reviewer 3's comment, we reviewed the cited PM+ papers, and confirm that the trial in Pakistan did in fact use "lay counsellors" with tertiary-level education. Therefore, we have removed these references, and instead cite only the PM+ trial from Kenya, which used community health workers with a minimum high-school level education. We thank the reviewer for identifying this oversight. We also acknowledge that in some cases, no tertiary education or formal certification may be required, but overqualified individuals are unlikely to be turned away during recruitment. In such a case, a few individuals with tertiary education or other certification may be operating among the otherwise "lay" staff. This may be acceptable, provided that the intervention does not require this certification, and that the majority of staff are still "lay workers" in the strictest sense. If we come across such cases over the course of the review, we will be certain to document and justify our decisions r.e. inclusion in the final review paper. However, in the interest of concision, we have decided not to expand upon this further in the text of the review protocol, as we have already reached the recommended maximum word limit, and at this point it would be speculative, anyway.

-p. 3, last line
We have reworded as per reviewer 3's suggestion, but also clarified that the Kenya study used community health workers with a minimum high school-level education.
-p. 11, limitations section, last line

We have revised to make clear that we are recommending future reviews focus on vulnerable populations that we were unable to include in our own review.

Thank you again and we look forward to hearing your final decision.