PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Supporting the use of theory in cross-country health services research: A participatory qualitative approach using Normalisation Process Theory as an example
AUTHORS	O'Donnell, Kate; Mair, Frances; Dowrick, CF; O'Reillyde Brun, Mary; de Brún, Tomas; Burns, Nicola; Lionis, Christos; Saridaki, Aristea; Papadakaki, Maria; van den Muijsenbergh, Maria; van Weel-Baumgarten, Evelyn; Gravenhorst, Katja; Cooper, Lucy; Princz, Christine; Teunissen, Erik; van den Driessen Mareeuw, Francine; Vlachadi, Maria; Spiegel, Wolfgang; Macfarlane, Anne

VERSION 1 - REVIEW

REVIEWER	Anne Lambert-Kerzner Eastern Colorado Health Care System USA
REVIEW RETURNED	04-Nov-2016

GENERAL COMMENTS	For the editors - I do not like this format for the review process as the answers to the above questions are not ye/no answers. For the authors - MS ID#: bmjopen-2016-014289 MS TITLE: Supporting the use of theory in cross-country health services research: Normalization Process Theory as an example. This paper is very interesting and will be a wonderful contribution to the D&I literature. I have a few suggestions that I believe may improve the impact of this work. Please see the attached document for this suggestions. Summary: This is well written manuscript describing in detail theories, challenges and application of the utilization of theories focusing on the Normalization Process Theory (NPT). And finally, the focus on a NPT training program. NPT is a theoretical framework that was used to guide the implementation of the RESTORE intervention and this paper describes the training process, implementation process, the evaluation, and recommendations for improving the training program. Major suggestion: The manuscript covers too many objectives, resulting in a complex reading. Therefore, I would suggest focusing on the paper on the description and reflection of utilization of the NPT. As you know, the NPT theory, its development and utilization have been previously published and could be easily referenced. This would enable you to expand on the details of the implementation of the NPT with the RESTORE intervention. The potential depth of the description of the

	implementation of the RESTORE intervention utilizing Minor comments: 1. Methods: a. A qualitative methodology was utilized to collect the data. Yet, there is no qualitative analytical methods described. Please include description of data collection, analyses, and assurance of rigor. b. Methods and results are integrated causing confusion. Please separate the design of the training components into the methods section. 2. Results: a. Methods and results are integrated causing confusion. Please separate the implementation of the NPT training components. b. It would be of interest to identify specific applications of the application of the specific NPT components to the G/Tis in each of the countries. c. Please separate the evaluation of the training components into the results section.

REVIEWER	Laura C. Leviton The Robert Wood Johnson Foundation United States of America
REVIEW RETURNED	30-Dec-2016

GENERAL COMMENTS	This paper has a lot of potential to be a valuable addition to the literature on applying theory to interventions in health services. The multi-site study, the fact that several disciplines were represented at each site, the involvement of practitioners and potential for cross-cultural applicability, are all extremely interesting. The selection of theory and what I understand about the specifics of intervention are all relevant. The supporting materials are illuminating. However, portions of the presentation were somewhat muddled. The paper would be greatly improved with a few specifics. Also, some typos need to be addressed, and perhaps a copy-editor with English as a first language could address this before acceptance. I caught some of these and list them, but certainly not all. The most important criticism relates to best practices in coding for qualitative data analysis and content analysis. This is easily fixed, as recommended below. The challenges the authors encountered in gaining consensus and consistency are hardly unique-- it's just that they are being applied to a challenging healthcare study. However, the authors claim in the discussion that they achieved "robust and consistent" coding (line 19, page 21) and that experiential training was "effective" (line 24, page 22) and again at line 3, page 24 that there was "a consistent view of constructs." However, no data are presented on effectiveness or consistency—in other words, what is the inter-rater reliability for that which is coded? This, most qualitative researchers would agree, is the criterion for whether qualitative coding training is effective, i.e. that the participants can recreate the same codes and agree on what is being observed. This can be handled by reporting kappa statistics-- not percent agreement, which is inflated by the number of "no" codings in a yes-no 2x2. Kappa controls for large numbers in one or more of the matrix cells). From there, it is easy to report an average kappa and a range across constructs or raters-- nobody expects perfect agreement, and information is readily available on what constitutes

	high, medium and low consistency. I strongly advise the above, not only for the sake of this paper, but also for the quality of other articles the team might generate. Noise and error in coding are bound to obscure important insights about the application of the NPT theory. We don't want that to happen-- we want to be able to detect the interesting patterns the team will find as the study unfolds. In fact, it would be interesting to see how kappa agreement varies by NPT construct, subconstruct, country, rater discipline, and the particular GT/I. It is not necessary to identify the country or rater discipline that indicates less than ideal agreement. Rather, it supports the article's contention that consistency was obtained across countries and disciplines. Where agreement is high or rather lower for the NPT constructs or subconstructs, that is extraordinarily valuable to know as well, and could really assist other qualitative investigators. Perhaps much of the above is for a second paper, but at a minimum, the average and range for the theory constructs, the countries involved, and the disciplines involved, are important. It may be that the investigators did not retain the individual ratings, or that they did not all rate the same instances. If so, it should be relatively easy to reconstruct coding for a sample of the qualitative data. That would be sufficient. It is also possible that the investigators were using a different criterion for consistency or effectiveness. If so, then that needs explanation. A related issue is that there are four rather complicated concepts to convey in this article: the participatory learning and action approach, the Normalisation Process Theory constructs and subconstructs, the specifics of the RESTORE study, and the specifics of guidance and training initiatives (GT/I). Agreement or lack of agreement about coding could be a result of clarity and training about the NPT subconstructs, the country/culture being coded, the preferences of individual participants nested within country, or the specific GT/I that was chosen for implementation and thus coding. To say that challenges arose because of the application of theory is thus complicated by these other concepts. Where were the difficulties located? This could be valuable information and not that hard to present. In general the concept of linearity vs reciprocal influences is not made clear. Examples from the study itself could explicate this. Table 2 is helpful but not entirely. Also what is described is fairly typical of qualitative analysis. Just because the study decided to train first on certain constructs does not mean that life stops. Another issue: it is not theory itself that is linear or recursive; it is the organisations' behavior and observables that are recursive. Indeed, NPT does have a linear or temporal sequence, which helps to explicate the relationships among constructs. This is important because there are a lot of vague references to linearity vs reciprocal factors in healthcare research these days, which is not particularly helpful. It should be clear that action can influence subsequent (temporal!) thinking at both individual and collective levels. But it is the observations which reflect this, not the theory itself. If the authors disagree with this, then fine, but they need to better explain what is meant.

	In the introduction to the participatory approach, it would be helpful to clarify that some of the coders were also implementers of the GT/Is. This was, at first highly confusing to me because one could apply a theory through specific intervention, or observe that which others did. It seems that the participatory approach involved both, correct? a few typos: in the abstract, "consisting [of?] 8 project..." "experience of clinical and non-clinical researchers were" – watch number unclear antecedent "it is not always applied consistently." Line 35 p7 Focussed—is this a British spelling, or a typo? appears twice in early pages.

VERSION 1 – AUTHOR RESPONSE

Reviewer 2.
Sample size - how many people were evaluated? Please make this more obvious in the methods/results. Did anyone not evaluate? All members of the team participated in the evaluation – 18 in total. We have amended both the abstract and text to make this clear.
Abstract now reads:
Primary outcome measures: Views of all research team members (n=18) assessed using qualitative evaluation methods, analysed thematically by the trainers after each session.
Methods, page 14 now reads:
Everyone present at the training days (generally all 18 members of the research team) participated in each evaluation; no one refused to participate.
Please also include a completed copy of the COREQ checklist This has been completed and enclosed as a supplemental file.

Reviewers' Comments to Author: Reviewer 1. We wish to thank Dr Lambert Kerzner for her very positive response and for her comments and suggestions which have strengthened and improved our paper.
Major suggestion:
The manuscript covers too many objectives, resulting in a complex reading. Therefore, I would suggest focusing on the paper on the description and reflection of utilization of the NPT. As you know, the NPT theory, its development and utilization have been previously published and could be easily referenced. This would enable you to expand on the details of the implementation of the NPT with the RESTORE intervention. Thank you for this suggestion to improve the clarity of our paper. We have reviewed and refined our aim and objectives at the end of the introduction to focus the aim on describing and reflecting on the utilisation of NPT (see page 7), removing some additional text. This also aligns the text more closely with the abstract.
The text now reads:
The aim of this paper is to describe and reflect on the process of designing training in the use of a theoretical framework in a multi-site cross-country research project.

We believe that some description of NPT is helpful to help researchers new to this theory understand our rationale for using it. However, in order to simplify our methodological description, we have moved our main description of NPT into Box 1, page 8 (see highlighted text). We have retained Table 1 which describes the specific application of NPT constructs and sub-constructs to the RESTORE project. Again, we feel that this is helpful for readers new to NPT.

Finally, we hope that the restructuring of the paper, as suggested by Reviewer 2, has also reduced the complexity of the paper. In particular, this includes ensuring there is no overlap between the

Methods and Results and sub-dividing both sections into Early, Mid and Late Project Training.

Minor comments:
1. Methods.
a. A qualitative methodology was utilized to collect the data. Yet, there is no qualitative analytical methods described. Please include description of data collection, analyses, and assurance of rigor.
There are two sources of qualitative data collection discussed in this paper. The first relates to the collection of data from the research team members themselves as part of the evaluation of the training. We describe the collection and analysis of these data in the Methods, Evaluation of NPT training content (page 14). These data included verbal feedback collected on the day using digital recorders, and written feedback in the form of short notes or commentaries from team members. We did not conduct in-depth qualitative analysis on these sources of data as they were, necessarily, brief. However, the material was reviewed by all four members of the training team and the evaluation feedback was reported back to the team for verification and agreement. We have added this detail into the Methods, page 14. The revised text now reads:

All the evaluation feedback was reviewed by the four members of the training team and the findings summarised into 'what worked', 'what didn't work' and 'what the team would like to do next'. The results of the evaluations were then summarised and presented back to the full team at the next face-to-face RESTORE Consortium meeting, providing the team with a further opportunity to comment on whether they believed all the key issues or suggestions regarding training had been captured and addressed.

The second source of qualitative data discussed in this paper is the fieldwork data. We have extended our description of the training we developed in the later stages of the project to support the research team conduct rigorous cross-country analysis (see Methods pages 12 -13). In doing this, we have extended the description of our approach in the Results (pages 16 - 17), adding a new text box to briefly describe the generation of this material (Box 2) and an additional table (Table 3). In Table 3, we use data from a coding exercise to demonstrate the coding from two in-country teams and the final coding, agreed by the RESTORE team. We hope this will demonstrate the rigor which we applied to training the team in cross-country qualitative data analysis, using NPT as a theoretical framework.

However, our focus in this paper is on describing the training to support analysis, not on the analysis of fieldwork data itself. Papers which contain a detailed description of the RESTORE fieldwork data analysis have recently been published. For example:
Lionis C et al. Engaging migrants and other stakeholders to improve communication in cross-cultural consultation in primary care: a theoretically informed participatory study. BMJ Open 2016: 6: e010822.
Teunissen E et al. Implementing guidelines and training initiatives to improve cross-cultural communication in primary care consultations: a qualitative participatory European study. International Journal for Equity in Health 2017 (accepted for publication).
b. Methods and results are integrated causing confusion. Please separate the design of the training components into the methods section. We have revised and re-organised the Methods and the Results to separate out methods and findings more clearly. This has also been more clearly explained in the Results by using the same sub-headings in relation to the timeframes of training as are used in the Methods (Early, Mid- and Late Project Training).

2. Results:

a. Methods and results are integrated causing confusion. Please separate the implementation of the NPT training components.
As described above, we have revised and re-organised the Methods and the Results to separate out

methods and findings more clearly. This has also been more clearly explained in the Results by using the same sub-headings in relation to the timeframes of training as are used in the Methods (Early, Mid- and Late Project Training).
b. It would be of interest to identify specific applications of the application of the specific NPT components to the G/Tis in each of the countries. We welcome this suggestion from Dr Kerzner, however we feel that to address that here would make the paper unduly long and unwieldy. We have recently reported on the use of NPT to appraise the implementability of the G/Tis identified in the early phase of the RESTORE work, which may be of interest.
de Brun et al. Guidelines and training initiatives that support communication in cross-cultural primary care settings: appraising their implementability using Normalization Process Theory. Family Practice 2015; 32: 420.
c. Please separate the evaluation of the training components into the results section. As previously described, we have now separated the reporting of the training evaluation into the results section. To increase the clarity of our reporting, we have also separated the results section into Early, Mid and Late Project Training.

Reviewers' Comments to Author: Reviewer 2.
We thank Dr Leviton for her supportive comments and we hope that our re-structuring and responses to her comments have improved the flow and clarity of our paper.
The most important criticism relates to best practices in coding for qualitative data analysis and content analysis. This is easily fixed, as recommended below. The challenges the authors encountered in gaining consensus and consistency are hardly unique-- it's just that they are being applied to a challenging healthcare study. However, the authors claim in the discussion that they achieved "robust and consistent" coding (line 19, page 21) and that experiential training was "effective" (line 24, page 22) and again at line 3, page 24 that there was "a consistent view of constructs." However, no data are presented on effectiveness or consistency—in other words, what is the inter-rater reliability for that which is coded? This, most qualitative researchers would agree, is the criterion for whether qualitative coding training is effective, i.e. that the participants can recreate the same codes and agree on what is being observed. This can be handled by reporting kappa statistics--not percent agreement, which is inflated by the number of "no" codings in a yes-no 2x2. Kappa controls for large numbers in one or more of the matrix cells). From there, it is easy to report an average kappa and a range across constructs or raters-- nobody expects perfect agreement, and information is readily available on what constitutes high, medium and low consistency. We have carefully considered this critique of our paper and we note that these comments relate in particular to our description of the later process of developing training in coding using NPT. We agree that we did not provide evidence to support our claim of achieving 'a consistent view of constructs' (previously on page 24, line 3). We also agree that it is important to demonstrate consistency in this process. However, we are less comfortable with the suggestion that we apply a quantitative construct such as Kappa scores to our qualitative coding. In our experience, this is a less common approach than demonstrating clear coding consistency through, for example, an audit trail of coding comparisons and reaching consensus within teams on the meaning and application of codes to data.

To address this, we have extended our description of our approach in the Results (pages 16 - 17) and added an additional table (Table 3). In that Table, we use data from a coding extract were we demonstrate the coding from two in-country teams and the final coding, agreed by the RESTORE team. We hope that this will address the issue to your satisfaction and demonstrate that we did indeed achieve a 'consistent view of constructs'.
I strongly advise the above, not only for the sake of this paper, but also for the quality of other articles the team might generate. Noise and error in coding are bound to obscure important insights about the application of the NPT theory. We don't want that to happen-- we want to be able to detect the interesting patterns the team will find as the study unfolds. As described above, we hope our

approach has dealt with your concerns. For additional information, we would also draw your attention to two recently published papers which report on our qualitative data analysis:
Lionis C et al. Engaging migrants and other stakeholders to improve communication in cross-cultural consultation in primary care: a theoretically informed participatory study. BMJ Open 2016: 6: e010822.
Teunissen E et al. Implementing guidelines and training initiatives to improve cross-cultural communication in primary care consultations: a qualitative participatory European study. International Journal for Equity in Health 2017 (accepted for publication).
In fact, it would be interesting to see how kappa agreement varies by NPT construct, subconstruct, country, rater discipline, and the particular GT/I. It is not necessary to identify the country or rater discipline that indicates less than ideal agreement. Rather, it supports the article's contention that consistency was obtained across countries and disciplines. Where agreement is high or rather lower for the NPT constructs or sub-constructs, that is extraordinarily valuable to know as well, and could really assist other qualitative investigators. Perhaps much of the above is for a second paper, but at a minimum, the average and range for the theory constructs, the countries involved, and the disciplines involved, are important.
This is a very interesting suggestion; however, we feel it is currently out with the scope of this paper. We will, however, give it careful consideration as we consider a further, methodological paper on the challenges and advantages of conducting cross-country qualitative fieldwork and analysis.
It may be that the investigators did not retain the individual ratings, or that they did not all rate the same instances. If so, it should be relatively easy to reconstruct coding for a sample of the qualitative data. That would be sufficient. It is also possible that the investigators were using a different criterion for consistency or effectiveness. If so, then that needs explanation.
As above, we hope our way of dealing with this issue addresses this point.
A related issue is that there are four rather complicated concepts to convey in this article: the participatory learning and action approach, the Normalisation Process Theory constructs and sub-constructs, the specifics of the RESTORE study, and the specifics of guidance and training initiatives (GT/I). Agreement or lack of agreement about coding could be a result of clarity and training about the NPT sub-constructs, the country/culture being coded, the preferences of individual participants nested within country, or the specific GT/I that was chosen for implementation and thus coding. To say that challenges arose because of the application of theory is thus complicated by these other concepts. Where were the difficulties located? This could be valuable information and not that hard to present.
We agree that there are four concepts to convey in our article; however the principle focus is on the NPT training and – by extension – the constructs and sub-constructs. The participatory learning and action approach, as applied to the RESTORE project, will be the focus of a future publication. We have retained our brief description of PLA in Box 1, along with three related references for those interested in further reading on this. Box 1 also contains more information about NPT itself.

We agree that agreement, or lack of agreement, about coding could be due to multiple factors. We have extended our description of the process of providing training in coding on pages 16 - 17, along with a new Table 3, which we hope clarifies the approach to training in relation to the cross-country qualitative analysis. We have also described this above.

An additional paragraph has been inserted into the Discussion section (page 24) to reflect on the influence of professional and cultural perspectives. This text reads:

A multi-disciplinary, multi-national team inevitably has differences in terms of understanding the process of qualitative research and the use of theory. Professional and cultural perspectives impact on both individual and collective comfort (both in terms of country and professional discipline) with the concept of using theory to inform the design and conduct of a largely qualitative, implementation study. For example, researchers used to a more inductive approach to data analysis were initially cautious of an approach that applied a theoretical framework to data analysis. The design of a robust

programme of training, which acknowledged and discussed these perspectives during the course of the training, was challenging but also allowed the team to reach a shared understanding of what the study was trying to achieve. The benefits of surfacing these tensions became apparent as the training moved to the process of data analysis.

In general, the concept of linearity vs reciprocal influences is not made clear. Examples from the study itself could explicate this. Table 2 is helpful but not entirely. Also what is described is fairly typical of qualitative analysis. Just because the study decided to train first on certain constructs does not mean that life stops. We are not entirely sure what is required by this comment. We have extended Table 2 to demonstrate how we compared qualitative analysis from each in-country RESTORE team and reached agreement over coding.

Figure 2 and the accompanying description, which also address this point, has been moved to nearer the beginning of the Methods section (see next comment). We have also extended our discussion of this issue (page 23). We hope that, taken together, we have now addressed this issue.

Another issue: it is not theory itself that is linear or recursive; it is the organisations' behavior and observables that are recursive. Indeed, NPT does have a linear or temporal sequence, which helps to explicate the relationships among constructs. This is important because there are a lot of vague references to linearity vs reciprocal factors in healthcare research these days, which is not particularly helpful. It should be clear that action can influence subsequent (temporal!) thinking at both individual and collective levels. But it is the observations which reflect this, not the theory itself. If the authors disagree with this, then fine, but they need to better explain what is meant. We have attempted to be more clear about our argument here, by altering and expanding the text.

First, we have moved the description of the alignment of training to each NPT construct to the Study Design section of the Methods (page 9). This provides an early acknowledgement that the application of the theory needs to have some temporal sequence. We also expanded our description of this issue in the ‘challenges’ section of our Results. This text now reads:

Although not entirely linear, the study was designed to broadly align to the four constructs of NPT (Figure 2). Stage 1 focused on familiarisation, first on the broad need to apply theory to RESTORE and then, with NPT itself. Stage 2 mapped to coherence and cognitive participation; Stage 3 mapped to collective action and reflexive monitoring. This structure then influenced the design of the training for the team, which is described below.

We agree that action can, indeed, influence subsequent thinking and we observed this during our training as well as in later fieldwork (not reported in this paper). We have articulated this more clearly with an example of one of the challenges that arose during training – text under Results, Challenges on page 23 now reads:

A final concern was whether construct application and data generation, in the field, was linear or whether there were ‘feedback loops’. For example, the research team considered the question of whether engaging in the work of implementing a G/TI could increase participants understanding or ‘coherence’ in relation to that G/TI. Training, therefore, continuously emphasised the lack of linearity in the process of applying theory to both data collection and analysis and encouraged the researchers to think through how this would affect data collection in the field.

In the introduction to the participatory approach, it would be helpful to clarify that some of the coders were also implementers of the GT/Is. This was, at first highly confusing to me because one could apply a theory through specific intervention, or observe that which others did. It seems that the participatory approach involved both, correct? We have amended the text on the RESTORE study design (page 9) to make this clearer. The text now reads:

In Stage 3, the selected G/TI was refined by local stakeholders supported by the in-country RESTORE team, implemented by the stakeholders and RESTORE team, monitored and, where necessary, further refined to improve the chances of sustaining it in routine practice.

Some typos need to be addressed, and perhaps a copy-editor with English as a first language could address this before acceptance. Thank you for picking up on inadvertent typos – these have been corrected. The first author is a native English speaker and UK English spelling has been used throughout.
a few typos: in the abstract, "consisting [of?] 8 project..."
"experience of clinical and non-clinical researchers were" – watch number
unclear antecedent "it is not always applied consistently." Line 35 p7
Focussed—is this a British spelling, or a typo? appears twice in early pages. We have tried to amend and correct typos and have amended 'focussed' to 'focused' throughout.

VERSION 2 – REVIEW

REVIEWER	Anne Lambert-Kerzner VHA USA
REVIEW RETURNED	03-Mar-2017

GENERAL COMMENTS	MS ID#: bmjopen-2016-014289.R1 MS TITLE: Supporting the use of theory in cross-country health services research: A participatory qualitative approach using Normalisation Process Theory as an example Summary: This is well written manuscript describing and reflecting on the process of designing and delivering a training program supporting the use of a theoretical framework, Normalization Process Theory (NPT), in a multi-site cross-country health services research study. NPT is a theoretical framework that was used to guide the implementation of the RESTORE intervention. This revised manuscript has succeeded in focusing on the three important components of the original paper: 1) utilization of a theoretical framework in the implementation of an evidence based intervention; 2) describing how it was accomplished; and 3) providing useful suggestions for training session in the utilization of theories, models, and frameworks in implementing interventions. These suggestions can easily be generalized across the use of theory in implementation science overall. Minor suggestions: 1) In the manuscript the use of the term and the description of "theories" in the background leaves one to wonder what is a "theoretical framework", which is the term that is used to describe NPT. I would suggest using the differential definitions of theories, models, and frameworks as described by Per Nilsen in his paper" Making sense of implementation theories, models and frameworks Implementation Science" (2015) 10:53 DOI 10.1186/s13012-015-0242-0. This, I believe, will better describe the differentiation between the three levels. The "grand, mid, and small" terms are not as clear and do not explain why you use the term "theoretical framework". 2) It would be very interesting if you could explore a little more about PLA and the integration of it with NPT in the Mid-Project Training (Months 13-24). Editing Very minor – page 14 of 93 line 56 sentence needs to be completed

	"process and the results are reported in……"

VERSION 2 – AUTHOR RESPONSE

Responses to Reviewer 1

Please find below a detailed description and explanation of the changes we made to our resubmitted manuscript. Where appropriate and to illustrate the changes made, the new text is shown here in italics, to clearly highlight the changes made. The marked-up copy of the manuscript had the changes highlighted in yellow.

Reviewer's Comments to Author: Reviewer 1. Again, we would like to thank Dr Lambert Kerzner for her very positive comments to our revised manuscript. We are pleased to address her last points, which have again improved the clarity of our paper.
Minor suggestions:
1) In the manuscript the use of the term and the description of "theories" in the background leaves one to wonder what is a "theoretical framework", which is the term that is used to describe NPT. I would suggest using the differential definitions of theories, models, and frameworks as described by Per Nilsen in his paper" Making sense of implementation theories, models and frameworks Implementation Science" (2015) 10:53 DOI 10.1186/s13012-015-0242-0. This, I believe, will better describe the differentiation between the three levels. The "grand, mid, and small" terms are not as clear and do not explain why you use the term "theoretical framework".

We acknowledge that our use of the term 'theoretical framework' may add confusion. Nilsen's paper describes NPT as an implementation theory; Carl May originally defined NPT as a mid-range theory. Thus, we have edited the paper to refer to NPT as a 'theory', as per both Nilsen and May.

2) It would be very interesting if you could explore a little more about PLA and the integration of it with NPT in the Mid-Project Training (Months 13-24).

While we agree that this is indeed interesting, we felt it was out with the scope of this paper. We have recently published a paper on the integration of PLA and NPT (Reference 31: de Brun et al. Learning from doing: the case for combining normalisation process theory and participatory learning and action research methodology for primary healthcare implementation research. BMC Health Services Research 2016;16:346. doi: 10.1186/s12913-016-1587-z) and others are currently in preparation.

Editing
Very minor – page 14 of 93 line 56 sentence needs to be completed "process and the results are reported in……"
This has been corrected.

VERSION 3 – REVIEW

REVIEWER	Anne Lambert-Kerzner, PhD, MSPH Eastern Colorado Healthcare System, VHA in Denver Co., USA
REVIEW RETURNED	24-Apr-2017

GENERAL COMMENTS	MS ID#: bmjopen-2016-014289.R1 MS TITLE: Supporting the use of theory in cross-country health services research: Normalization Process Theory as an example. Summary: This is well written manuscript describing and reflecting on the process of designing and delivering a training program supporting the use of a theory, Normalization Process Theory (NPT),

	in a multi-site cross-country health services research study. NPT is a theory that was used to guide the implementation of the RESTORE intervention. This revised manuscript has succeeded in focusing on the three important components of the original paper: 1) utilization of a theory in the implementation of an evidence based intervention; 2) describing how it was accomplished; and 3) providing useful suggestions for training session in the utilization of theories, models, and frameworks in implementing interventions. These suggestions can easily be generalized across the use of theory in implementation science overall.

VERSION 3 – AUTHOR RESPONSE

Many thanks for your recent correspondence and the final request from the editors with respect to our manuscript. I have now added a short section, titled 'Limitations of the study' to the Discussion on page 26.