PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Fidelity assessment of nurse-led non-pharmacological package of care for knee pain in the package development phase of a feasibility randomised controlled trial based in secondary-care: a mixed methods study
AUTHORS	Nomikos, Polykarpos; Hall, Michelle; Fuller, Amy; Millar, Bonnie; Ogollah, Reuben; Valdes, A; Doherty, Michael; Walsh, David; dasNair, Roshan; Abhishek, A

VERSION 1 – REVIEW

REVIEWER	Walton, Holly University College London, Department of Applied Health Research
REVIEW RETURNED	17-Nov-2020

GENERAL COMMENTS	Summary Overall, I really enjoyed reading this manuscript and it provides an interesting insight into the delivery of a nurse-led non-pharmacological intervention for those with knee pain. The authors have used mixed-methods to assess fidelity and explore factors influencing fidelity of delivery. I think that this manuscript will be of interest to those working in the fields of intervention fidelity and also those interested in knee pain. The manuscript is well written and presented throughout. I have added some comments on each section for the authors' consideration. A couple of general comments are also listed below: 1. In some places throughout the manuscript (abstract and introduction), it isn't clear whether the participants mentioned are the intervention participants or those included in the fidelity assessment (or both), e.g.: □ Abstract: May be worth specifying whether the participants/inclusion criteria mentioned in the abstract are those who took part in the intervention or whether those are the sample that were chosen for the fidelity assessment □ Abstract: Also unclear in the methods when you say that all sessions were video-recorded. All sessions in the whole intervention, or were a proportion of sessions chosen for the fidelity assessment? □ Methods: The participants and recruitment section provides a really clear description of the intervention participants (though it may be helpful to say how many participants received the intervention in total), but I would like a little more about the sampling of the video/self-report/interview aspects for the fidelity study specifically. E.g. were all participants that took part in the intervention recorded?

	How many recordings in total (across how many sessions)? Did the provider complete the self-report checklists for every participant? How many self-report checklists were completed in total for each session? How was the provider chosen for delivery of the programme? 2. Minor point but there are a couple of grammatical errors throughout the manuscript. □ 'semi-structured interviews were…' (abstract) □ Time constraints (Introduction – line 22 p4) Abstract: The abstract is concise and clearly written. 3. May be helpful to provide the design in the abstract (mixed methods) 4. Were the fidelity checklists developed specifically for the study? 5. Number of checklists completed (self-report and video ratings) and number of interviews conducted may be helpful 6. Might be helpful to clarify that the findings presented in the last sentence of the results are from the interviews. 7. It may be helpful to provide specific implications from your study Introduction: The introduction provides a clear introduction to osteoarthritis and the need for non-pharmacological introductions, introduction to what fidelity is, and then introduces the current study 8. The introduction may benefit from further context-setting in relation to fidelity methods including: □ how fidelity is assessed (e.g. from your paragraph in the discussion) □ Why mixed methods evaluations of fidelity are needed (e.g. from your paragraph in the discussion) 9. The introduction could specify why the study is needed a little more clearly (i.e. what is the gap for the study) – this would then lead nicely into the aims of the study Methods: 10. Study design – further justification of why mixed methods may be helpful here & also the sequence of the mixed-methods and how the different elements interacted (this is particularly key since you've done so much work integrating the data!) 11. How did the PPI members support the design of the study? Further information may be helpful 12. I wonder if it may be helpful to slightly revise the structure of the methods section around: a) the intervention itself b) the fidelity assessment and c) the qualitative study 13. Further information about how the fidelity checklists were developed would be helpful – i.e what steps were followed? 14. Did the researchers follow any guidelines to code rate for fidelity against the checklist? If so, further information is needed 15. How was the topic guide for the interview developed? 16. 'An additional interview was conducted ' - At what time points? 17. You have mentioned that the fidelity assessment followed that of previous studies – but further details on which of their methods you adopted may be helpful 18. Data analysis – qualitative - further information about the general inductive approach may be needed for those unfamiliar with this 19. Data analysis – fidelity assessment – further information on how the overall fidelity % reported in table 4 was calculated 20. Ethics – given that the study includes findings from one provider delivering the intervention, I wondered if it may be worth providing more details within the ethics section around anonymity in terms of publishing quotes/quantitative findings from one provider.

	Results: The results section is clearly written and is very interesting to read. The integration of findings from the quantitative and qualitative elements is really clear. 21. I like that you have described the findings across different areas of the intervention components, and across sessions 1-4 – I would be interested to know if the fidelity scores varied across different participants? (given that it is the same provider) Discussion: The discussion provides a clear summary of the key findings, strengths and limitations, future research and implications. 22. I wonder if your paragraphs (2/3/4) on fidelity methods and reasons for using mixed methods in the discussion may be more appropriate in the introduction – and then in the discussion you could focus on what your study has done to extend previous research / and how your findings support previous research e.g. for the fidelity assessment - differences between self-report and video-ratings, and also perhaps for the qualitative findings too – e.g. do your findings support previous research on barriers/facilitators to fidelity?). □ On that last point - I personally found it really interesting reading your qualitative findings as there are some similar findings (despite being different interventions & condition) with the barriers/facilitators we found in my study on barriers/facilitators of fidelity of delivery for a complex intervention for people with dementia (https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-020-01006-x). I'm wondering if this could perhaps indicate some key overlapping themes that may limit/facilitate fidelity of delivery despite different types of intervention (e.g. perhaps there are also some overlapping themes preventing delivery in Toomey's (2017) mixed methods study and other studies too 23. Was engagement of the participants considered in this study? If not, it may be worth mentioning this in the strengths/limitations Tables: Really clear tables – very easy to follow 24. Table 1 –May be worth specifying that these are the intended components that should be delivered in each session (at first glance I thought this was a results table showing which components were delivered during the intervention) 25. Table 2 and Table 3 – please outline how many fidelity checklists each table is based on (overall and for each session). 26. Table 4 – does the overall fidelity percentage refer to the nurse self-report or the video observations? Would be interested to see whether it converges with both the self-report and the video observations.

REVIEWER	Okuno, Yuji Edogawa Hosp, MSK intervention center
REVIEW RETURNED	15-Dec-2020

GENERAL COMMENTS	This novel mixed methods study used a combination of techniques to assess treatment fidelity. This study addresses an important aspect in the proper implementation of non-pharmacological treatment of OA. The fact that only one nurse participated in the study is a weak point. The introduction, methods, results, and discussion are very well organized and well written throughout.

	The research methods and the results are reasonable and I do not see any problems. It would have been better if you could mention some of the problems in using this method for future RCTs.

REVIEWER	Mallett, R Sheffield Hallam University
REVIEW RETURNED	15-Dec-2020

GENERAL COMMENTS	I feel there are significant gaps in reporting or insufficient criteria that may call into question the ethical basis of intervention justification in a group that were not identified to be patients with OA where OA guidelines have been the sole focus of treatment. Without exclusion criteria worrisome or non-musculoskeletal pathology may also be present in this cohort. As participants in what appears to be a larger IMHW study this may have occurred but is not reported. 1. Is the research question or study objective clearly defined? Yes I certainly feel it is defined however I have significant concerns regarding the disconnect between the application of OA guidelines to a cohort that appear to be poorly identified by inclusion and exclusion criteria. OA is a radiological finding. The inclusion criteria without exclusion criteria leads to questions that red flag pathology may have existed and was not fully assessed. Without this safeguards patient safety is at risk. This may have occurred and is not reported. I feel reporting should clarify this and also be more explicit regarding assessment prior to intervention. The below inclusion criteria: The inclusion criteria were: age>40 years, ability to read and write in English, knee pain present for longer than three months, pain in or around the knee on most days of the previous month, and at least moderate pain in two of the five domains of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain scale would not identify patients with OA. A variety of peripheral neurogenic, myogenic and soft tissue origins to pain alongside possible referred pain and possibly vascular or centrally sensitised pain would be present in this group. In the absence of further assessment and diagnosis I feel it should be reported that all possible underlying pathologies were treated in a generic fashion by applying interventions from OA guidelines that are likely in the majority of the above examples have some benefit. 2. Is the abstract accurate, balanced and complete? Page 3, 13: I would disagree this is individualised intervention when no specific patient centred care has occurred and all patients have been treated with the same pathology as the focus of their care. Detail is lacking how interventions have been individualized. For example: Without strength examination it is difficult to prescribe patient centred strengthening regimes. 3. Is the study design appropriate to answer the research question?

	4. Are the methods described sufficiently to allow the study to be repeated? "The nurse explained aerobic and strengthening exercises and advised each participant on individualised regimens. If required, weight-loss advice was provided." As stated above detail is lacking how regimes were individualized. How was weight loss advice decided to be required? 5. Are research ethics (e.g. participant consent, ethics approval) addressed appropriately? Please see comment 1 and also no ethical consideration is described regarding recruitment, confidentiality, right to withdraw etc etc through both arms of the mixed methods. These need addressing. Perhaps it is the written statement however clarification of what the practitioner and the independent reviewer are blinded to. There appears to be discrepancy from the abstract that suggests the reviewer is fully blinded whereas in the methods it states they are blinded to the nurse ratings. 11. Are the discussion and conclusions justified by the results The conclusions do address the question of whether intervention has been applied as intended. Considering it is reported that the package was individualised a dichotomous completed measure seems a blunt tool to evaluate whether certain aspects of treatment was applied as intended ie. individualised. Measurement of sets and reps of exercises prescribed from the results of the initial assessment would be required for this. There is also little consideration of patient adherence or compliance. The intervention was explained or offered however little discussion of whether this was adhered too once away from the practitioner is documented. Partially completed is present on the assessment form. How is this quantified? 12. Are the study limitations discussed adequately? Please see the points above. 15. Is the standard of written English acceptable for publication? Largely yes, there are one or two sentences that would benefit from structural changes. For example: Page 23, 11: The study was of people with knee pain but none had evidence of inflammatory arthritis and in this age group, OA is the main cause of chronic usage related knee pain. Further comments • Pg 3, 51: OA not identified as study focus. Inclusion criteria states knee pain for 3 months this could be many other MSK or neurological conditions. • Poor inclusion criteria - ideally the cohort needs to be identified as a population with OA to apply the NICE guidelines for OA to their best effect and intended purpose. These are common interventions that are unlikely to be detrimental to anyone with joint pain however the cohort may have non MSK masquerades with knowing exclusion criteria.

	• Ethical considerations that participants were given an education booklet on OA could be argued implicitly suggests a 'OA' diagnosis to the patient when the inclusion criteria does not led to identifying this patient cohort. I would suggest this is not ethical sound. • Page 22, line 16. OA may have been found to be the most current underlying pathology however this does not been everyone in this age bracket should be treated as though they have the pathology. Further discussion and justification of applying these guidelines to this cohort would aid understanding the research process and transferability of general guidelines to broadly defined joint pains. • PAge 24, line 24 - there is no evidence these are arthritis patients. I would suggest defining patients by an radiological finding is not helpful and a more patient centred approach is where recent evidence and literature has led this field of research. • Page 25, 15 - Sorry I do not understand what you mean by categories in this statement.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1.

Reviewer Name: Dr. Holly Walton

Institution and Country: University College London, United Kingdom

Comment 1*:* In some places throughout the manuscript (abstract and introduction), it isn't clear whether the participants mentioned are the intervention participants or those included in the fidelity assessment (or both).

Response: Fidelity of delivery of intervention was assessed for all study participants who received the intervention. This is now stated in the abstract:

"Each intervention session with every participant was video recorded and formed part of fidelity assessment" (page2, lines 16-17).

Comment 2*:* May be worth specifying whether the participants/inclusion criteria mentioned in the abstract are those who took part in the intervention or whether those are the sample that were chosen for the fidelity assessment.

Response: All participants who received the intervention were included in the fidelity assessment. This is now specified in the abstract:

"Nurse self-report and assessor video rating scores for all 62 treatment sessions were included in fidelity assessment" (pages2-3 lines 23, 1). Please see response to reviewer comment 1.

Comment 3: Also unclear in the methods when you say that all sessions were video-recorded. All sessions in the whole intervention, or were a proportion of sessions chosen for the fidelity assessment?

Response: All sessions in the whole intervention were video recorded. More specifically: "*Eighteen participants received the non-pharmacological intervention and all (n=62) sessions were video-recorded" (page10, lines 3-4).*

Methods

Comment 4: The participants and recruitment section provides a really clear description of the intervention participants (though it may be helpful to say how many participants received the intervention in total). I would like a little more about the sampling of the video/self-report/interview aspects for the fidelity study specifically. E.g. were all participants that took part in the intervention recorded? How many recordings in total (across how many sessions)?

Response:

This is now stated both in the Methods and in Results sections, specifically:

*"Eighteen participants received the non-pharmacological intervention and all (n=62) sessions were video-recorded" (page10, lines 3-4). **"**After every session with the participant, the nurse completed the fidelity checklist. Sixty-two checklists, 18 for session 1, 16 for session 2 and 14 each for sessions 3 and 4 were completed" (page10, lines 4-6).*

Comment 5*:* Did the provider complete the self-report checklists for every participant? How many self-report checklists were completed in total for each session? How was the provider chosen for delivery of the programme?

Response:

Please also see response to reviewer comment 4. *"After every session with the participant, the nurse completed the fidelity checklist." "Sixty-two checklists, 18 for session 1, 16 for session 2 and 14 each for sessions 3 and 4 were completed" (page10, lines 4-6).*

"The nurse delivering the intervention was working as a research nurse previously and did not have prior knowledge of musculoskeletal diseases, had not worked in rheumatology or allied specialities such as orthopaedics, rehabilitation or sports medicine, and never delivered treatments for arthritis" (pages7-8, lines 23-25, 1-2).

***Comment 6*:** Minor point but there are a couple of grammatical errors throughout the manuscript.
'semi-structured interviews were…' (Abstract)
Time constraints (Introduction)

Response: We thank the reviewer for noticing the minor grammatical errors. We have rectified them.

Abstract: *"Two semi-structured interviews were conducted with the research nurse" (page2, lines 20-21).*

Introduction: *"However, this can be difficult for general practitioners (GPs) to deliver for several reasons such as time constraints and, core non-pharmacological treatments are under-utilised" (page4, lines 11-13).*

Abstract

***Comment 7*:** The abstract is concise and clearly written. May be helpful to provide the design in the abstract (mixed methods).

Response: Added. *"Study design: mixed methods study" (page2, line 5).*

Comment 8: Were the fidelity checklists developed specifically for the study?

Response: The fidelity checklists were developed specifically for this intervention, based on a previous validated checklist for complex interventions. More specifically:

"Our study is based on a fidelity checklist that has been previously validated in complex interventions delivered in a research setting. We tailored the checklist according to the intervention and further refined it" (page24, lines 17-19).

In the Methods section, we stated that we followed the relevant five-step methodology to develop the checklist. More specifically:

"The fidelity checklist was iteratively developed using a five-step methodology. These were: reviewing previous measures, analysing intervention components and developing an intervention framework (intervention manual), developing the fidelity checklist, obtaining feedback about the content and wording of checklist and piloting and refining the checklist to assess and improve reliability" (page9, lines 17-21).

Comment 9: Number of checklists completed (self-report and video ratings) and number of interviews conducted may be helpful

Response: These have now been added in the Abstract and the Methods, specifically:

"Nurse self-report and assessor video rating scores for all 62 treatment sessions were included in the fidelity assessment" (pages2-3, lines 23, 1).

"Sixty-two checklists, 18 for session 1, 16 for session 2 and 14 each for sessions 3 and 4 were completed" (page10, lines5-6).

"Two semi-structured interviews were conducted with the research nurse" (page2, lines 20-21).

Comment 10: Might be helpful to clarify that the findings presented in the last sentence of the results are from the interviews.

Response: Thank you for the comment. We have now clarified this in the Results, specifically:

"The nurse reported difficulty advising on thermal treatments, footwear and walking aids, and did not feel confident negotiating achievable and realistic goals with participants" (page3, lines 4-6).

Comment 11: It may be helpful to provide specific implications from your study

*Response: **"**Future research should assess intervention fidelity in a routine clinical setting, and examine its clinical and cost-effectiveness" (page3, lines8-10).*

Introduction:

Comment 12: The introduction provides a clear introduction to osteoarthritis and the need for non-pharmacological introductions, introduction to what fidelity is, and then introduces the current study. I wonder if your paragraphs (2/3/4) on fidelity methods and reasons for using mixed methods in the discussion may be more appropriate in the introduction. The introduction may benefit from further context-setting in relation to fidelity methods including: How fidelity is assessed (e.g. from your paragraph in the discussion? Why mixed methods evaluations of fidelity are needed (e.g. from your paragraph in the discussion). The introduction could specify why the study is needed a little more clearly (i.e. what is the gap for the study) – this would then lead nicely into the aims of the study.

Response: We thank the reviewer for the suggestions, and the relevant paragraph from the discussion has been moved to introduction.

"There are several methods to assess treatment fidelity, including direct observation, patient self-report questionnaire, provider self-report checklist, and indirect observation using audio or video-recordings…….Video recordings were chosen as this is less intrusive than direct observation and provide an opportunity to assess reliability" (page5, lines 1-15).

"Medical Research Council guidelines for developing and evaluating complex interventions have highlighted the importance of conducting process evaluation…..For this reason, a mixed methods approach was utilised" (page5, 16-23).

We hope that we have addressed the comments appropriately and the connection between the gap and the aims of the study is now improved.

Methods:

Comment 13: Study design – further justification of why mixed methods may be helpful here & also the sequence of the mixed-methods and how the different elements interacted (this is particularly key since you've done so much work integrating the data!)

Response: We provided further justification of why mixed methods is helpful on that particular design. We also provided information how the different elements interacted. More specifically:

"This form of mixed methods approach was used to produce additional insights of the issue at hand. In this design, qualitative and quantitative data are collected and complementary results arise from the use of different methods. In the current study, the quantitative data informed the collection of qualitative data and a convergence approach was followed" (page7, lines 3-7).

Comment 14: How did the PPI members support the design of the study? Further information may be helpful

Response: They advised that video recording of treatment sessions would be acceptable to participants (page8, lines 5-6).

Comment 15: I wonder if it may be helpful to slightly revise the structure of the methods section around: a) the intervention itself b) the fidelity assessment and c) the qualitative study

Response: We thank the reviewer for the suggestion. Structure of the methods are revised accordingly. As:

a) Intervention (page8)

b) Fidelity assessment and analysis (pages9-11)

c) The qualitative study (pages11-12).

Comment 16: Further information about how the fidelity checklists were developed would be helpful – i.e what steps were followed? Did the researchers follow any guidelines to code rate for fidelity against the checklist? If so, further information is needed

Response: We thank the reviewer for the comment. We followed published strategies to code rate fidelity and referenced the paper:

"The fidelity checklist was developed a priori and comprised eight components, each with specific tasks: materials; introduction; assessment; education; exercise; weight loss; advice on adjunctive treatments; and review and planning"(page9, lines11-14).

"The fidelity checklist was iteratively developed using a five-step methodology" (page9, lines 17-18). Also, please see response to reviewer comment 8.

"The scoring criteria of the fidelity checklist followed that of previous published strategies for assessing fidelity in RCTs of complex interventions" (page10, lines 1-2).

Comment 17: How was the topic guide for the interview developed?

Response: "The interview guide contained open-ended questions developed by the study team, which included a rheumatologist (AA), physiotherapists (MH, PAN), psychologist (RdN), and qualitative researcher (AF)"(page11, lines 6-8).

Comment 18: 'An additional interview was conducted ' - At what time points?

Response: "*an additional interview was conducted 45 weeks later to capture any salient points raised from the initial quantitative and qualitative data collected" (page11, lines 12-14).*

Comment 19: You have mentioned that the fidelity assessment followed that of previous studies – but further details on which of their methods you adopted may be helpful

Response: "The study followed the National Institutes of Health Behaviour Change Consortium (NIHBCC) guidelines for fidelity assessment" (page9, lines 10-11).

"A combination of provider self-report and independent assessed video recording was utilised in the current study to provide an in-depth fidelity assessment" (page5, lines 11-13).

Comment 20: Data analysis – qualitative - further information about the general inductive approach may be needed for those unfamiliar with this

Response: The steps for the general inductive approach are provided:

"Transcripts were analysed following the principles of the general inductive approach. The latter is a simple straightforward approach, which is used to derive findings from raw qualitative data, condense them into a brief summary format, and link the research objectives with the summary findings" (page12, 3-6).

*Comment 21***:** Data analysis – fidelity assessment – further information on how the overall fidelity % reported in table 4 was calculated

Response: We thank the reviewer for the comment. Table 4 presents fidelity as median (%). We report on how median fidelity is assessed and added the Interquartile Ranges within table 4:

"To obtain fidelity score for a component of the intervention, individual scores for each task within the component were added and divided by the maximum possible score for that component and converted to a percentage" (page10, lines 14-17).

"Median fidelity scores (%) and IQR were calculated for the entire intervention, per participant, per session and per component of the intervention" (page10, lines19, 20)

Comment 22: Ethics – given that the study includes findings from one provider delivering the intervention, I wondered if it may be worth providing more details within the ethics section around anonymity in terms of publishing quotes/quantitative findings from one provider.

Response: An ethical approval section was added and a section around the anonymity of the nurse details.

"The study received ethical approval by the East Midlands-Derby Research Ethics Committee (REC) (18/EM/0288)" (page9, lines 4-5).

***"**We have not provided demographic or other details in order to protect the anonymity of the individual nurse" (page11, lines 17-19).*

Results:

The results section is clearly written and is very interesting to read. The integration of findings from the quantitative and qualitative elements is really clear. I like that you have described the findings across different areas of the intervention components, and across sessions 1-4.

Comment 23: I would be interested to know if the fidelity scores varied across different participants? (given that it is the same provider)

Response: We thank the reviewer for this suggestion. We further analysed the data and provided an extra Table 3:

The fidelity scores that were assessed using video-recordings across participants (page 17, Table 3)

"For individual participants, overall fidelity across the four sessions ranged from 75% to 100% indicating that for most patients the intervention was delivered as intended (Table 3)" (page14, lines 8-10).

Discussion:

Comment 24: In the discussion you could focus on what your study has done to extend previous research / and how your findings support previous research e.g. for the fidelity assessment -

differences between self-report and video-ratings, and perhaps for the qualitative findings too – e.g. do your findings support previous research on barriers/facilitators to fidelity?

Response: We appreciated the feedback from the reviewer. We added how the findings support previous research for fidelity assessment-differences between self-report and video-ratings:

"*In our study, the research nurse rated themselves higher than the independent rating using the video recordings consistent with previous studies" (page25, lines 19-21).*

In addition, for the qualitative findings:

"From the interview transcripts, factors that influenced fidelity of delivery are identified. The nurse was less confident to identify appropriate patient goals and prescribe exercise in the first few sessions, but this improved thereafter. This is not a barrier per se, but suggests that some further training and additional support for nurses in this new role would be needed to ensure fidelity at the start of the study. The nurse was able to draw on her previous experience working with other patient groups to discuss and assess complex issues. Nurse's previous experience assessing patients, therefore, facilitated fidelity of delivery. Although the fidelity for education appeared to be lower in the first session this was because the nurse recognised and responded that participants were being given a lot of information" (pages24-25, lines 22-24, 1-7).

Previous studies using mixed methods have explored factors that influenced fidelity and found good fidelity of delivery of a physiotherapist-led complex package of care for chronic low-back pain and OA. They report on the factors that influenced fidelity on three levels: provider, participant and programme (page 25, lines 13-16).

Comment 25: On that last point - I personally found it really interesting reading your qualitative findings as there are some similar findings (despite being different interventions & condition) with the barriers/facilitators we found in my study on barriers/facilitators of fidelity of delivery for a complex intervention for people with dementia (https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-020-01006-x). I'm wondering if this could perhaps indicate some key overlapping themes that may limit/facilitate fidelity of delivery despite different types of intervention (e.g. perhaps there are also some overlapping themes preventing delivery in Toomey's (2017) mixed methods study and other studies too

Response: We thank the reviewer for this comment. We considered and followed appropriate guidelines to develop the fidelity checklist to assess fidelity of delivery. We added the relevant sections to support previous research on barriers to fidelity.

"Similar findings on barriers and facilitators to deliver the intervention have been identified in a complex intervention for people with dementia and chronic low back pain. In fact, Walton et al extended over the factors that influenced fidelity of delivery reported by Toomey et al and recognised

that knowledge, providers' attributes, ease of adaptation of the intervention in relation to participants' needs influenced fidelity. Based on the findings, it was challenging to address adaptation and determine the appropriate balance between fidelity and adaptation in this study. This may indicate some key overlapping themes that may limit fidelity of delivery despite the different types of intervention and conditions" (pages 25-26, lines 21-25, 1-5).

Comment 26: Was engagement of the participants considered in this study? If not, it may be worth mentioning this in the strengths/limitations

Response: We did not consider engagement of the participants in our study and this was added as a limitation: *"We did not consider to capture engagement of the participants in the study" (page26, line 16).*

Tables Comments.

Really clear tables – very easy to follow

Comment 27: Table 1 –May be worth specifying that these are the intended components that should be delivered in each session (at first glance I thought this was a results table showing which components were delivered during the intervention)

Response: We thank the reviewer for the comment. We excluded Table 1 from the manuscript as it has already been published in our protocol.

Comment 28: Table 2 and Table 3 – please outline how many fidelity checklists each table is based on (overall and for each session).

Response: Tables 2 and 3 have now become Tables 1 and 2. We added the number of sessions (checklists) completed during each session and overall. We specified them with an asterisk * *(pages15, 16).*

Comment 29: Table 4 – does the overall fidelity percentage refer to the nurse self-report or the video observations? Would be interested to see whether it converges with both the self-report and the video observations.

Response: The overall fidelity percentage in table 4 refers to the video observations. We compared it with the video observations, because we believed that it would provide a more objective insight into what fidelity would be after integrating the findings. Looking at the data from table 2 (self-report findings), it seems that qualitative findings would converge with all the intervention components apart from the adjunctive treatments.

Reviewer 2

Comment 1: The fact that only one nurse participated in the study is a weak point.

Response: We thank the reviewer for the comments. We highlighted the fact that only one nurse participated in our study as a limitation:

"Strengths and limitations: A single nurse was involved in delivery of the intervention" (page3, line 21).

"There are a number of limitations to this study. A key caveat is that only one nurse was involved in delivery of the intervention" (page26, lines 6-7)

The introduction, methods, results, and discussion are very well organized and well written throughout.

The research methods and the results are reasonable and I do not see any problems.

Comment 2: It would have been better if you could mention some of the problems in using this method for future RCTs.

Response: We thank the reviewer for raising this point. We report on the problems in using this method for future RCTs:

"In a larger trial, there would be more nurses to deliver the intervention across multiple sites, which increases the likelihood of variation in fidelity" (page26, lines 7-9).

"Follow-up training sessions and support during the start of the feasibility when nurses are first delivering the intervention may be helpful in order to improve confidence and delivery" (page27, lines 15-17).

Reviewer 3

Comment 1: I have significant concerns regarding the disconnect between the application of OA guidelines to a cohort that appear to be poorly identified by inclusion and exclusion criteria. OA is a radiological finding. The inclusion criteria without exclusion criteria leads to questions that red flag pathology may have existed and was not fully assessed. Without this safeguards patient safety is at

risk. This may have occurred and is not reported. I feel reporting should clarify this and also be more explicit regarding assessment prior to intervention. The below inclusion criteria:

The inclusion criteria were: age>40 years, ability to read and write in English, knee pain present for longer than three months, pain in or around the knee on most days of the previous month, and at least moderate pain in two of the five domains of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain scale would not identify patients with OA. A variety of peripheral neurogenic, myogenic and soft tissue origins to pain alongside possible referred pain and possibly vascular or centrally sensitised pain would be present in this group. In the absence of further assessment and diagnosis I feel it should be reported that all possible underlying pathologies were treated in a generic fashion by applying interventions from OA guidelines that are likely in the majority of the above examples have some benefit.

Response: Thank you for this concern. We can confirm that the research nurse undertook a full history and performed musculoskeletal assessment to confirm that the participant had knee osteoarthritis. The NICE guidelines for osteoarthritis advise not to use plain radiography for the diagnosis of this condition in the presence of typical symptoms in the at-risk age group. We can confirm that we used the same strategy in this study. The research nurse was trained in these by Dr Hall, and could consult the CI (Abhishek) if the clinical features were not consistent with OA. Thus, we are satisfied that we have adhered to NICE guidelines for diagnosing osteoarthritis. We also want to highlight to the reviewer that there is poor correlation between structural changes and symptoms in plain radiography and typically symptoms precede radiographic changes by several years.

"At the first visit, the nurse took a medical history, examined the knee joints and explained to the participant that they had knee pain due to OA. Investigations and radiographs were not undertaken as per NICE guidelines" (page8, lines 13-14).

"In the presence of activity related joint pain, no or minimal morning stiffness, and age >= 45 years, a clinical diagnosis of OA may be reached without the need of investigations (e.g. blood tests or radiography) as per the National Institute for Health and Care Excellence (NICE) guidelines" (page4, lines 5-9).

"The Chief Investigator (AA) was available for advice if a clinical diagnosis of OA could not be reached. In that case, the participant would be deemed ineligible for the study" (page8, lines15-17).

"Where the nurse identified difficulties in delivering the intervention as intended, she was able to seek additional advice and training from MH. This experience has allowed us to further improve the nurse training programme for use in the feasibility RCT" (page25, lines 10-12). .

However, we agree with the reviewer that periarticular pathologies are common in people with knee osteoarthritis and people with and without these were treated with exercise, weight loss and education.

Comment 2: Page 3, 13: I would disagree this is individualised intervention when no specific patient centred care has occurred and all patients have been treated with the same pathology as the focus of their care. Detail is lacking how interventions have been individualized. For example: Without strength examination it is difficult to prescribe patient centred strengthening regimes.

Response: The intervention has been described in our published protocol (Hall, et al (2020) BMJ Open **10**(9): e037760)

It consists of a holistic assessment and delivery of core components (patient information, exercise and weight loss if required) that are individualised to the participant. The participant's understanding of their condition, previous treatments, goals and preferences are discussed and are used to individualise the intervention.

For the exercise components, a knee examination including assessment of muscle strength, flexibility and function was conducted by the nurse. Exercises were selected on the basis of the assessment and the participants own goals. Physical activity goals and aerobic exercise were individualised after taking into account participant preferences and what they feel can be realistically achieved.

Comment 3: "The nurse explained aerobic and strengthening exercises and advised each participant on individualised regimens. If required, weight-loss advice was provided." As stated above detail is lacking how regimes were individualized. How was weight loss advice decided to be required?

Response: Weight loss advice was provided to any participants who had a BMI over 25kg/m^2.The NHS BMI calculator was used to determine how much weight a participants would need to lose to have a healthy BMI and estimate how long it might take (based on a 500 calorie reduction per day). https://www.nhs.uk/livewell/healthy-weight/bmi-calculator/.

Beliefs about eating, physical activity and weight and their knee pain were explored and any previous experiences of losing weight discussed. No single approach was promoted by the nurse, participants were signposted to NHS weight loss plan but were also free to use any method they prefer eg a commercial weight loss plan/group or other online weight loss apps. An evidence based weight loss goal of 10% was discussed with participants but for most an initial short term goal of 5% was set.

Comment 4: Are research ethics (e.g. participant consent, ethics approval) addressed appropriately? Please see comment 1 and also no ethical consideration is described regarding recruitment, confidentiality, right to withdraw etc etc through both arms of the mixed methods. These need addressing. Perhaps it is the written statement however clarification of what the practitioner and the independent reviewer are blinded to.

Response: We thank the reviewer for the comments. We have provided details about ethical approval.

"Ethical approval: The study received ethical approval by the East Midlands-Derby Research Ethics Committee (REC) (18/EM/0288) (page9, lines4-5).

Relevant sections on consent have been added:

"All study participants including the research nurse gave their written informed consent prior to treatment delivery, including the consent to video record the sessions. Participants had the right to pause or stop the video recording at any point without giving any reasons" (page9, lines 6-9).

& More information regarding consent and withdrawal can be found on the published protocol: (Hall, et al (2020) BMJ Open **10**(9): e037760):

"Participants will be free to withdraw at any time if they desire to do so, or at the discretion of the chief investigator. In the event of withdrawal, any data collected up until that point will be kept and potentially included in any analyses"

Comment 5: There appears to be discrepancy from the abstract that suggests the reviewer is fully blinded whereas in the methods it states they are blinded to the nurse ratings.

Response: We thank the reviewer for this comment. We have rectified the relevant problematic sections.

Abstract:

"Fidelity checklists were completed by the research nurse after each session and by an independent researcher, after viewing the video-recordings blinded to nurse ratings" (page2, lines 17-19).

Methods:

"Blinded to the nurse ratings, the video-recording of every session was independently reviewed and rated by PAN" (page10, lines 6-7).

Comment 6: Are the discussion and conclusions justified by the results

The conclusions do address the question of whether intervention has been applied as intended. Considering it is reported that the package was individualised a dichotomous completed measure seems a blunt tool to evaluate whether certain aspects of treatment was applied as intended ie.

individualised. Measurement of sets and reps of exercises prescribed from the results of the initial assessment would be required for this. There is also little consideration of patient adherence or compliance. The intervention was explained or offered however little discussion of whether this was adhered too once away from the practitioner is documented. Partially completed is present on the assessment form. How is this quantified?

Response:

We agree that assessing fidelity of an individualised and complex intervention such as this is difficult. The checklist evaluated whether the key components needed to individualise the intervention were addressed by the nurse, for example exploring participants' health beliefs including concerns, expectations and knowledge is required to ensure advice given is individualised. Assessment of muscle strength and function and discussion of participant goals would be required in order to individualise exercise selection and prescription. To overcome the limitations of the checklist the qualitative interviews allowed us to explore whether participants perceived that the intervention was individualised to them.

"Whilst checklists can be helpful in determining whether an intervention has been delivered they do not allow for or capture the flexibility that is required when tailoring an intervention to the individual" (page27, lines 4-6).

"One week after the final session, the nurse took part in a semi-structured interview conducted by PAN (PhD student) and AF (trained qualitative researcher)" (page11, lines 4-8).

"The guide covered the nurse's view on their training, confidence in and experience of delivering the individual components of the non-pharmacological intervention, perceived barriers to delivering it as planned, and opportunities to improve the non-pharmacological package of care" (page11, lines 8-12).

"A second interview with the nurse was conducted to capture any salient points not discussed during the first interview" (page26, lines14-15).

With respect to exercise prescription, aerobic activity was prescribed using the FITT principles (frequency, intensity, type and time) and muscle strengthening was prescribed in line with guidelines from the American College of Sports Medicine. We did not include the prescription of the exercise in the checklist and agree this is a limitation of the fidelity assessment.

Adherence to the intervention is not the same as fidelity of delivery. We did not measure adherence directly in the phase of the study although it was addressed in the qualitative interviews.

"We did not consider to capture engagement of the participants in the study" (page26, line 16).

Where items on the fidelity checklist were assessed as partially completed, the assessor felt that the item could have been addressed further, for example if a participant indicated that activity pacing was problematic but this was not fully explored by the nurse with recommendations on how it could be improved.

We report on how partially completed was scored:

"The responses of the fidelity checklist were categorical and rated as completed, partially completed, not completed, or not applicable. Partially completed scores were given for any task that was not delivered to the full extent in the context of that particular consultation" (pages9-10, lines 21-24, 1).

Comment 7: Is the standard of written English acceptable for publication?

Largely yes, there are one or two sentences that would benefit from structural changes. For example: Page 23, 11: The study was of people with knee pain but none had evidence of inflammatory arthritis and in this age group, OA is the main cause of chronic usage related knee pain.

Response: As outlined above, if the nurse thought participants had symptoms of inflammatory arthritis they could consult with the CI, a rheumatologist. We have therefore left this sentence unchanged.

Comment 8: OA not identified as study focus. Inclusion criteria states knee pain for 3 months this could be many other MSK or neurological conditions. Poor inclusion criteria - ideally the cohort needs to be identified as a population with OA to apply the NICE guidelines for OA to their best effect and intended purpose. These are common interventions that are unlikely to be detrimental to anyone with joint pain however the cohort may have non MSK masquerades with knowing exclusion criteria.

Response: Please see response to reviewer comment 1. We can confirm that the same strategy was used when diagnosing OA as has been recommended in the NICE guidelines.

Comment 9: Ethical considerations that participants were given an education booklet on OA could be argued implicitly suggests a 'OA' diagnosis to the patient when the inclusion criteria does not led to identifying this patient cohort. I would suggest this is not ethical sound.

Response: As outlined in response to reviewer comment 1, based on participant assessment nurse informed them that they had knee OA before proceeding further. NICE guidelines do not mandate a plain radiograph to diagnose OA in the presence of typical symptoms and in the at-risk population.

Comment 10: Page 22, line 16. OA may have been found to be the most current underlying pathology however this does not been everyone in this age bracket should be treated as though they have the pathology. Further discussion and justification of applying these guidelines to this cohort would aid understanding the research process and transferability of general guidelines to broadly defined joint pains.

Response: The participants were carefully assessed by a trained research nurse to reach a diagnosis of OA and to exclude other forms of arthritis or conditions that may present with knee pain. We have developed a package of care that is deliverable in primary-care in the UK and uses the diagnostic strategy and management planning as outlined in the NICE guidelines.

Comment 11: PAge 24, line 24 - there is no evidence these are arthritis patients. I would suggest defining patients by an radiological finding is not helpful and a more patient centred approach is where recent evidence and literature has led this field of research.

Response: Please see response to reviewer comments 1, 8, 9

Comment 12: Page 25, 15 - Sorry I do not understand what you mean by categories in this statement.

Response: Thank you for pointing this out. We have changed categories to components in this sentence and elsewhere

VERSION 2 – REVIEW

REVIEWER	Walton, Holly University College London, Department of Applied Health Research
REVIEW RETURNED	15-Mar-2021

GENERAL COMMENTS	Review of: Assessing intervention fidelity of a nurse-led non-pharmacological package of care for knee pain Manuscript ID: bmjopen-2020-045242 The authors have provided clear and appropriate responses to reviewers' comments and have amended the manuscript accordingly. I enjoyed reading the revised manuscript. I have included some very minor amendments below: - Introduction page 5 line 2/3 – minor grammatical changes - patient self-report questionnaires and provider self-report checklists - Page 5 line 9 – the references after the word indeed – it's not clear which part of the sentence these refer to – maybe move to the end of the sentence? - Page 25 line 21-23 – maybe amend the sentence slightly to make it clear that Walton et al and Toomey et al are two different studies – e.g. 'have been identified in two complex interventions (one for people with dementia and one for people with chronic low back pain)'.

REVIEWER	Mallett, R Sheffield Hallam University
REVIEW RETURNED	23-Mar-2021

GENERAL COMMENTS	Dear Authors, I thank you for your detailed responses and read your comments with interest. Many minor revisions have significantly improved the transparency of the methodology, diagnosis in this population, consent and ethical considerations. I do apologise if my comments pointed towards the need for plain radiographs. This was never the intention and as you state not indicated by the NICE guidelines. It was more the exclusion of other pathology that was was required to leave a hypothesised diagnosis in this at risk group. Thank you for consideration of the tricky elements surrounding adherence to recommended management plans.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1.
Reviewer Name: Dr. Holly Walton
Institution and Country: University College London, United Kingdom

Comment 1: Introduction page 5 line 2/3 – minor grammatical changes - patient self-report questionnaires and provider self-report checklists

Response: We thank the reviewer for suggesting minor grammatical changes. We rectified them.

“There are several methods to assess treatment fidelity, including direct observation, patient self-report questionnaires and provider self-report checklists, and indirect observation using audio or video-recordings”.

Comment 2: Page 5 line 9 – the references after the word indeed – it’s not clear which part of the sentence these refer to – maybe move to the end of the sentence?

Response: We thank the reviewer for noticing that the references were misplaced. References are now moved to the end of the sentence as suggested.

“Indeed, it has been shown previously that assessing fidelity using independently rated recordings and provider self-report checklist is feasible and acceptable.” (References).

Comment 3: Page 25 line 21-23 – maybe amend the sentence slightly to make it clear that Walton et al and Toomey et al are two different studies – e.g. ‘have been identified in two complex interventions (one for people with dementia and one for people with chronic low back pain)’.

Response: We thank the reviewer for providing this suggestion. We changed the sentence accordingly and made it clear that Walton et al and Toomey et al are two different studies.

“Similar findings on barriers and facilitators have been identified in two complex interventions, one for people with dementia and one for people with chronic low back pain."