PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Feasibility and potential effectiveness of a non-pharmacological multidisciplinary care programme for persons with generalised osteoarthritis: a randomised multiple-baseline single-case study.
AUTHORS	Hoogeboom, Thomas; Kwakkenbos, Linda; Rietveld, Leonie; den Broeder, Alfons; de Bie, Rob; van den Ende, Cornelia

VERSION 1 - REVIEW

REVIEWER	Nadine Foster Prof of Musculoskeletal Health in Primary Care Arthritis Research UK Primary Care Centre Keele University Keele UK I have no conflicts of interest
REVIEW RETURNED	04-May-2012

THE STUDY	There are several aspects of the study that are a little unclear and remain un-justified. Specifically - why was a multidisciplinary intervention selected as the intervention - there is no clear justification for this given in the paper. Was this intervention already available or was it developed specifically for this research study? - similarly why was the intervention group-based? No justification is given for this. - the intervention summary box is useful but highlights that it really mostly comprised information-giving/education. Yet we know from previous research in clinical conditions that education is a rather weak intervention to change behaviour. Thus the authors need to justify the components of the intervention more clearly. Also specifically what was the activity programme - did it focus on best evidence to date in focusing on strengthening and aerobic exercise? The papers says 'focus on quality of movement' - what is meant by this and why was this the focus rather than strengthening exercise (for which there is most evidence for effectiveness in OA)? Also the authors state the intervention was tailored but do not provide any information on how it was tailored? Would some specific examples be useful. It must be challenging to truly tailor a group-based intervention? - the exclusion criteria 'if therapists suspected high levels of distress' is unclear and unjustified. How was it assessed? What is meant by it? - page 7 states that adherence was measured but the paper never explains how -page 10 - why was 20% change deemed clinically relevant? Whilst it seems reasonable, other research has shown a need fo 30% or more. Again, what is the justification for 20%?

	- the team selected a feasibility study or different design to the future hoped for RCT. Why was a pilot RCT not carried out if the ultimate plan was to inform a main RCT? - overall the sample size, even for single case research, is small (only 4 of 5 provided data) and ultimately the study is based on only one small group that received the intervention as a group of OA patients. - I didn't quite follow the authors argument that the research shows they should not do a main RCT, I would have thought that the research shows clearly that the content and process of delivery of the intervention needs significant re-thinking but that ultimately a future main RCT would still be the right way to move forward to test its effectiveness. - reference 6 is missing some details - reference 14 refers to a RCT protocol - I was confused by this. Is this protocol for a different RCT with a different intervention? - Table 1 - how was education level determined? Table 2 seems a bit meaningless with only p values; could average data summary statistics be added? Table 3 needs a fuller footnote explaining all abbreviations Figures - label phases a, b and A'

REVIEWER	Dr Bhasker Amatya Department of Rehabilitation Medicine, Royal Melbourne Hospital, Victoria, Australia.
REVIEW RETURNED	23-May-2012

THE STUDY	1. Research question is vague and confusing, needs to be shortened and indication of patient population is missing. For example, the term “preliminary effectiveness” is not explicable. I would suggest the review of the title for e.g. ‘Feasibility and effectiveness of a non-pharmacological MD care programme for persons with GOA: a randomised multiple-baseline single-case study”. 2. Not sure if this is the appropriate design, as main aim of the study as anticipated by the authors are feasibility and effectiveness of the MD programme. I am not sure how feasibility can be assessed using this design, as measuring the dependent variable prior to administering treatment is an important aspect of this type of study. The effectiveness can sure be measured to some extend, however, the authors did not explain how the severity of the problem is quantified with measurement of the pain in a baseline period before treatment is introduced (as it seems pain scores in a VAS scale seems to be low threshold at baseline in majority of patients- in 3 out of 4). In addition, the A-B-A design assumes that when treatment is withdrawn, the condition would return to at least nearly what it was before the treatment began. However, with the multidisciplinary interventions the authors suggested usually we would expect to have a more lasting effect for longer-time, requiring longer follow-up. Furthermore, confounding variables (medication, age, disease duration etc.) is usually not possible with this design, and there is possibility that these confounding factors other than the treatment could have influenced the result. 3. Definition of the GOA needs to be elaborated (Methods section, first paragraph: line 11-14) 4. Not consistency with the primary and secondary measures through out the abstract and text. e.g. in abstract the authors

	indicates that feasibility as a primary outcome and effectiveness as secondary. However, in text in multiple occasions pain and self-efficacy are indicated as primary outcomes. 5. Interventions: not consistent through out the text. Please note non-pharmacological and multidisciplinary (MD) intervention are two broad terms and have diverse definition. For e.g. non-pharmacological intervention range from exercise/physical modalities to orthotics and education, where as MD intervention might be non-pharmacological and pharmacological, as well as non-pharmacological programme only provided by more than 2 disciplines. Needs to define the intervention in more details and needs to be consistent. 6. The authors statement in key message (first dot point and in introduction): '…no-studies are available that evaluate non-pharmacological care in GOA' seems not accurate, as there are lots of systematic reviews and studies evaluating these interventions in OA, which can be generalised to the GOA. 7. Joint pain is the cardinal sign of any OA including GOA, not comorbidities as stated in Introduction, should this be 'generalised joint-pain' instead? Please review.
RESULTS & CONCLUSIONS	1. Practical (or clinical) significance of the findings is not clear as it seems the intervention has not made a meaningful difference in the well-being of the participant. However, authors comment in Discussion section stating that '…current intervention does not warrant further evaluation in RCT' is arguable. As this might be due to the study design itself as the intervention was provide in a group and not tailored to patient needs and goal oriented. 2. It is well recognised that the sample consisting of a single subject engaged in a particular intervention provided by a particular individual is challenging, particularly in this study, due to the broad nature of the intervention. Usually, direct replication, systematic replication, and clinical replication is required for generalizability of the results from single-subject designs. Trialling the intervention using other study design with more participants and a control group would be ideal. 3. Introducing the patient recruitment procedure at the beginning might be helpful to the reader. How many were asked to participate, how many refused. 4. Feasibility of the program is arguable as the median expectation of participant prior to the programme (md=7) and perceived therapy effects (md=6). 5. The authors fail to set a cut-off score for both pain and self-efficacy, which would have aid to inspect for changes in level (magnitude) or reductions in variability. 6. Risk that evaluator bias and/or demand characteristics of the patients (e.g. not motivated) needs to be addressed as this might have influence the results. 7. The discussion section should include, What is the take home message for readers? 8. Figures need modifications: needs to indicate the A-B-A' in all figures.
GENERAL COMMENTS	Interesting study, with interesting and challanging study design. I think this manuscript needs revision, as there is lack of consistency regarding the aims of the study and the results reported in the abstract. Also, there is a lack of consistency in the aims reported in the abstract and the ones reported in the manuscript.

VERSION 1 – AUTHOR RESPONSE

We like to thank both Nadine Foster and Bhasker Amatya for their remarks and comments on our manuscript. We strongly believe the article has improved considerably regarding its quality, clarity and reproducibility and that we were able to incorporate the suggestions successfully.
The following list shows in detail how we dealt with each of the problems that the reviewers noted.We want to point out to the reviewers that the references made to page and line numbers comply with the marked manuscript.

Reviewer 1: Nadine Foster

Reviewer 1's remark 1:
Why was a multidisciplinary intervention selected as the intervention - there is no clear justification for this given in the paper. Was this intervention already available or was it developed specifically for this research study?

Comment to reviewer 1's remark 1:
We have added our justification for selecting a multidisciplinary intervention to the article as well as the statement that the intervention was specifically developed for this study (please see page 5, lines 13-16).

Reviewer 1's remark 2:
Similarly why was the intervention group-based? No justification is given for this.

Comment to reviewer 1's remark 2:
We have added a justification on why the intervention was group-based to the manuscript, please see page 10, line 3-4.

Reviewer 1's remark 3:
The intervention summary box is useful but highlights that it really mostly comprised information-giving/education. Yet we know from previous research in clinical conditions that education is a rather weak intervention to change behaviour. Thus the authors need to justify the components of the intervention more clearly. Also specifically what was the activity programme - did it focus on best evidence to date in focusing on strengthening and aerobic exercise? The papers says 'focus on quality of movement' - what is meant by this and why was this the focus rather than strengthening exercise (for which there is most evidence for effectiveness in OA)? Also the authors state the intervention was tailored but do not provide any information on how it was tailored? Would some specific examples be useful. It must be challenging to truly tailor a group-based intervention?

Comment to reviewer 1's remark 3:
We initially kept this section brief due to fact that we included the Pat-plot in our manuscript, however we agree with the reviewer's remarks that some of the aspects are too briefly described and need further clarification. Therefore we rewrote most of the 'Intervention'-paragraph, and added information addressing the reviewer's concerns. Please see paragraph Intervention (page 10-11).

Reviewer 1's remark 4:
The exclusion criteria 'if therapists suspected high levels of distress' is unclear and unjustified. How was it assessed? What is meant by it?

Comment to reviewer 1's remark 4:

If the therapists believed the patients with high distress levels would negatively impact the group process, patients were excluded from the group-based programme and offered an individual intervention. This clarification was included in the manuscript (page 7, line 12). Since the additional value of the use of validated questionnaires as a screening instrument for this purpose has not yet been proven, these were not incorporated in the present study and judgments were based on clinical experience.

Reviewer 1's remark 5:
Page 7 states that adherence was measured but the paper never explains how.

Comment to reviewer 1's remark 5:
Adherence to the multi-disciplinary therapy was determined by determining the number of no-shows to the actual therapy. We added this to page 8, line 13.

Reviewer 1's remark 6:
Page 10 - why was 20% change deemed clinically relevant? Whilst it seems reasonable, other research has shown a need for 30% or more. Again, what is the justification for 20%?

Comment to reviewer 1's remark 6:
We wanted to comply with the cut-off's used by the OARSI-responders criteria [Escobar A, Gonzalez M, Quintana JM, Vrotsou K, Bilbao A, Herrera-Espiñeira C, Garcia-Perez L, Aizpuru F, Sarasqueta C. Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement. Identification of cut-off values. Osteoarthritis Cartilage. 2012 Feb;20(2):87-92. Epub 2011 Nov 20.]. We added the reference to the manuscript (page 12, line 8).

Reviewer 1's remark 7:
The team selected a feasibility study or different design to the future hoped for RCT. Why was a pilot RCT not carried out if the ultimate plan was to inform a main RCT?

Comment to reviewer 1's remark 7:
As discussed in the manuscript, both the pilot RCT and single case study provide useful data for preparing a large RCT regarding the feasibility of the intervention as well as preliminary information on its effectiveness. As we were more interested in the feasibility of the intervention, rather than for example issues with randomization or sampling we decided to choose the design of the single-case study.

Reviewer 1's remark 8:
Overall the sample size, even for single case research, is small (only 4 of 5 provided data) and ultimately the study is based on only one small group that received the intervention as a group of OA patients.

Comment to reviewer 1's remark 8:
We agree with the reviewer that the sample is fairly small and have discussed this throughout the paper (see for example the Article Summary - Strengths and limitations of this study). We believe, however, that despite the small sample size, the present study provides useful information on the intervention and points for improvement. Furthermore, the study underlines the importance of piloting interventions and therefore serves as an example for other researchers.

Reviewer 1's remark 9:
I didn't quite follow the authors argument that the research shows they should not do a main RCT, I would have thought that the research shows clearly that the content and process of delivery of the intervention needs significant re-thinking but that ultimately a future main RCT would still be the right way to move forward to test its effectiveness.

Comment to reviewer 1's remark 9:
We agree with the reviewer that, although points of improvement were found for the present intervention, a RCT should ultimately be conducted to further study the effectiveness of multi-disciplinary interventions for GOA. What we meant with our conclusion was, that the intervention as described in this paper should not be evaluated in a randomized clinical trial, as it will most likely result in disappointing outcomes and there is room for improvements. To make this clearer, we have adjusted our manuscript's conclusion. Please see page 19, line 16-17.

Reviewer 1's remark 10:
Reference 6 is missing some details

Comment to reviewer 1's remark 10:
Thank you, the paper has just now been published and can be referred to in more detail.

Reviewer 1's remark 11:
Reference 14 refers to a RCT protocol - I was confused by this. Is this protocol for a different RCT with a different intervention?

Comment to reviewer 1's remark 11:
This reference describes the protocol for a RCT, in which a different multidisciplinary intervention is tested than described in this paper.

Reviewer 1's remark 12:
Table 1 - how was education level determined?

Comment to reviewer 1's remark 12:
We have added the meaning of the education levels Low, Medium and High education to the text (page 9, line 8-10).

Reviewer 1's remark 13:
Table 2 seems a bit meaningless with only p-values; could average data summary statistics be added?

Comment to reviewer 1's remark 13:
We agree with the reviewer that Table 2 seemed a bit meaningless the way it was presented in the manuscript. However, we do not think adding average data summary statistics would be a solution, as these data (n=4) will add very little information. Therefore, we decided to remove this table and implement its content in the manuscript's text (please see page 14, lines 18-20).

Reviewer 1's remark 14:

Table 3 needs a fuller footnote explaining all abbreviations

Comment to reviewer 1's remark 14:
We have clarified Table 3 by expanding the footnote. (Note: Table 3 is now Table 2)

Reviewer 1's remark 15:
Figures - label phases a, b and A'

Comment to reviewer 1's remark 15:
We have updated our figures (and their legends) according to the reviewer's recommendation.
Reviewer 2: Bhasker Amatya

Reviewer 2's remark 1:
Research question is vague and confusing, needs to be shortened and indication of patient population is missing. For example, the term "preliminary effectiveness" is not explicable. I would suggest the review of the title for e.g. 'Feasibility and effectiveness of a non-pharmacological MD care programme for persons with GOA: a randomised multiple-baseline single-case study".

Comment to reviewer 2's remark 1:
We have adjusted the title along the recommendations of the reviewer. In addition, we removed the word "preliminary" from our manuscript and replaced it by the term "potential".

Reviewer 2's remark 2a:
Not sure if this is the appropriate design, as main aim of the study as anticipated by the authors are feasibility and effectiveness of the MD programme. I am not sure how feasibility can be assessed using this design, as measuring the dependent variable prior to administering treatment is an important aspect of this type of study.

Comment to reviewer 2's remark 2a:
We understand the reviewer's concerns, but we do not fully agree with them. We believe the single-case design can be used to investigate the feasibility of an intervention, as long as the limitations of the single-case design are taken into account. For example, this design does not allow researchers to test issues regarding randomization or to determine the number of eligible non-volunteers (we have discussed this in the paper). On the other hand, it does allow to study the feasibility of the intervention itself and to determine whether evaluation of the program in a large randomized clinical trial would be worthwhile, or that further adjustments to its content are warranted.

Reviewer 2's remark 2b:
The effectiveness can sure be measured to some extent, however, the authors did not explain how the severity of the problem is quantified with measurement of the pain in a baseline period before treatment is introduced (as it seems pain scores in a VAS scale seems to be low threshold at baseline in majority of patients- in 3 out of 4).

Comment to reviewer 2's remark 2b:
In this study patients had to report functional disabilities in their daily living (HAQ-DI score of 0.5 or higher); this is part of the GOA definition which is now added to the manuscript. However, we do agree with the reviewer that additional thresholds for pain and/or self-efficacy levels would have been of value in selecting patients eligible for the intervention. We therefore have addressed this point of concern in our limitations paragraph in the discussion (page 17, lines 18-21).

Reviewer 2's remark 2c:
In addition, the A-B-A design assumes that when treatment is withdrawn, the condition would return to at least nearly what it was before the treatment began. However, with the multidisciplinary interventions the authors suggested usually we would expect to have a more lasting effect for longer-time, requiring longer follow-up. Furthermore, confounding variables (medication, age, disease duration etc.) is usually not possible with this design, and there is possibility that these confounding factors other than the treatment could have influenced the result.

Comment to reviewer 2's remark 2c:
Indeed, the A-B-A' design assumes that the therapy effect should vanish after the B-phase. We however specifically chose for this design-type as this allowed us to more clearly distinguish between the treatment phase (B-period) and the post-treatment phase (A'), as many single case studies describe the study effectiveness during the intervention phase (B-period). So, even though the design type might imply that we expected the therapy results to disappear, we explicitly describe that we expect the effect to be superior to the initial phase (A-period) (see page 11). We could describe the whole article as if we have used an A-B design, however that way we would have to eliminate a whole number of interesting data points.
The point raised by the reviewer that the study design does not allow researchers to correct for potentially confounding factors is true. As we find it important to point this out to the reader, we have stated this in our discussion section (page 18, line 21-23).

Reviewer 2's remark 3:
Definition of the GOA needs to be elaborated (Methods section, first paragraph: line 11-14)

Comment to reviewer 2's remark 3:
We have now stated the definition of GOA in the paper (please see page 7, lines 6-8).

Reviewer 2's remark 4:
Not consistency with the primary and secondary measures throughout the abstract and text. e.g. in abstract the authors indicates that feasibility as a primary outcome and effectiveness as secondary. However, in text in multiple occasions pain and self-efficacy are indicated as primary outcomes.

Comment to reviewer 2's remark 4:
This is the result of unclear writing. Feasibility is the primary outcome, but pain and self-efficacy are the main outcomes of interest in our research question on the effectiveness of the intervention. We have changed these vague statements throughout the manuscript (please see page 9, line 4-5 and page 14, line 3 & 18).

Reviewer 2's remark 5:
Interventions: not consistent throughout the text. Please note non-pharmacological and multidisciplinary (MD) intervention are two broad terms and have diverse definition. For e.g. non-pharmacological intervention range from exercise/physical modalities to orthotics and education, where as MD intervention might be non-pharmacological and pharmacological, as well as non-pharmacological programme only provided by more than 2 disciplines. Needs to define the intervention in more details and needs to be consistent.

Comment to reviewer 2's remark 5

We have changed these inconsistencies throughout the manuscript, now labeling our intervention as a “non-pharmacological, multidisciplinary intervention”. Moreover, we have described our intervention into more detail on page 10 and 11.

Reviewer 2’s remark 6:
The authors statement in key message (first dot point and in introduction): ‘…no-studies are available that evaluate non-pharmacological care in GOA’ seems not accurate, as there are lots of systematic reviews and studies evaluating these interventions in OA, which can be generalised to the GOA.

Comment to reviewer 2’s remark 6:
We agree with the reviewer that there are systematic reviews of interventions in OA, but as far as we know, none of those reviews actually provide data on persons with GOA. For this reason, the National Institute for Health and Clinical Excellence (NICE) included a statement in their OA guideline that trials in specifically people with GOA are absent and need to be performed. Therefore, we believe our statement is accurate.

Reviewer 2’s remark 7:
Joint pain is the cardinal sign of any OA including GOA, not comorbidities as stated in Introduction, should this be ‘generalised joint-pain’ instead? Please review.

Comment to reviewer 2’s remark 7:
We have changed the phrasing of joint-pain comorbidities into multiple joint involvement, which is more accurate in this context.

Reviewer 2’s remark 8:
Practical (or clinical) significance of the findings is not clear as it seems the intervention has not made a meaningful difference in the well-being of the participant. However, authors comment in Discussion section stating that ‘…current intervention does not warrant further evaluation in RCT’ is arguable. As this might be due to the study design itself as the intervention was provide in a group and not tailored to patient needs and goal oriented.

Comment to reviewer 2’s remark 8:
Even though the intervention was group-based, we did tailor the different aspects of the intervention to the individual health needs by means of goal setting. We have made our intervention more clear and reproducible by describing it into more detail in the Intervention paragraph (Page 10 and 11).

Reviewer 2’s remark 9:
It is well recognised that the sample consisting of a single subject engaged in a particular intervention provided by a particular individual is challenging, particularly in this study, due to the broad nature of the intervention. Usually, direct replication, systematic replication, and clinical replication is required for generalizability of the results from single-subject designs. Trialling the intervention using other study design with more participants and a control group would be ideal.

Comment to reviewer 2’s remark 9:
We agree that research on therapy options in this group of patients should not be aborted due to the negative results found in this study. However the studied intervention in its current form needs some rethinking before we re-evaluate it in scientific study. We changed our conclusion (page 19, line 13-17) accordingly to make this point clear to the reader.

Reviewer 2's remark 10:
Introducing the patient recruitment procedure at the beginning might be helpful to the reader. How many were asked to participate, how many refused.

Comment to reviewer 2's remark 10:
We have added this information to the paper, please see page 13, lines 3-5.

Reviewer 2's remark 11:
Feasibility of the program is arguable as the median expectation of participant prior to the programme (md=7) and perceived therapy effects (md=6).

Comment to reviewer 2's remark 11:
We agree with the reviewer. We have discussed this in the second paragraph of our discussion and one of the key messages states this as well.

Reviewer 2's remark 12:
The authors fail to set a cut-off score for both pain and self-efficacy, which would have aid to inspect for changes in level (magnitude) or reductions in variability.

Comment to reviewer 2's remark 12:
We agree with the reviewer on this point. As stated earlier (Reviewer 2's remark 2b), we have addressed this issue in our limitation section of the discussion (please see page 17, lines 18-21).

Reviewer 2's remark 13:
Risk that evaluator bias and/or demand characteristics of the patients (e.g. not motivated) needs to be addressed as this might have influence the results.

Comment to reviewer 2's remark 13:
It is not likely that an evaluator bias occurred in our study, as most of the measures were completed at the participants' home. Also, the pre- and postintervention questionnaires were send out by mail. However the impact of patients' characteristics on the (lack of) treatment effects is indeed important. We have added this statement to the discussion section (please see page 18, lines 21-23).

Reviewer 2's remark 14:
The discussion section should include, What is the take home message for readers?

Comment to reviewer 2's remark 14:
We agree a take home message is important, however we believe the take home message is described pretty clearly in the Article Summary – Key Messages section of the paper (please see page 3).

Reviewer 2's remark 15:
Figures need modifications: needs to indicate the A-B-A' in all figures.

Comment to reviewer 2's remark 15:
We have updated our figures (and their legends) according to the reviewer's recommendation.