PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Using an internet intervention to support self-management of low back pain in primary care: Findings from a randomised controlled feasibility trial (SupportBack)
AUTHORS	Geraghty, Adam; Stanford, Rosie; Stuart, Beth; Little, Paul; Roberts, Lisa; Foster, Nadine; Hill, Jonathan; Hay, Elaine; Turner, David; Malakan, Wansida; Leigh, Linda; Yardley, Lucy

VERSION 1 – REVIEW

REVIEWER	Rod Taylor University of Exeter Medical School,UK
REVIEW RETURNED	24-Mar-2017

GENERAL COMMENTS	The paper describes the design and results of an RCT to assess the feasibility of full trial of impact of an internet intervention to support self-management of low back pain in primary care with a primary functional outcome and 3-months. The study is well designed and generally well presented. However, there are some issues: Major The authors undertake an inferential between group comparison of outcomes. This analysis is not specified in the protocol. Moreover the way this is reported in abstract (in addition to usual improved by 0.6 RMDQ points…") gives inappropriate emphasis to outcome results of what is a pilot trial that is not powered to detect a difference in outcomes. The authors should either remove this result from abstract or at least present the 95% CI around this mean estimate. The authors do draw attention to the exploratory nature of this analysis in discussion. Minor The abstract needs to clarify the 3 groups who took part in the study – suggest use approach of abstract of protocol paper.

REVIEWER	Daniel Cury Ribeiro University of Otago - New Zealand
REVIEW RETURNED	03-May-2017

GENERAL COMMENTS	General comments Thank you for inviting me to review this feasibility trial. This study aimed to assess the feasibility of a RCT of an internet-based intervention for LBP delivered in addition to (1) usual care or (2) usual care + telephone support with a physiotherapist; and compared usual care.

	A total of 87 participants took part in the study, and primary outcomes were related to the feasibility of the trial (i.e. recruitment rate, adherence, and retention at follow-up). Exploratory analyses were conducted on clinical outcomes at 3 months follow-up. Results suggest that the full trial is feasible. This is a really well-written, and nicely conducted feasibility trial. I have some minor comments for the authors, and hope these help to improve the clarity of the manuscript. Specific comments Abstract: Apologies, but I found the aim description a little confusing. By reading the aims of the study, I was not clear whether you designed the trial for assessing “internet + usual care” versus “usual care”, and “internet + usual care + telephone support” versus “usual care”. Is ‘stand-alone’ intervention referring to internet + usual care? While reading the methods section of the abstract, I could not figure out what exactly the three arms were. I suggest you to revise it, to enhance clarity. Page 2, line 31: The term “trial time frame” is very generic, and does not give an idea of how long you recruited participants for… I think you could state what the length of period was. Strengths and limitations of the study Page 3, line 16: you did not mention physical activity measurements at the abstract anywhere. I am aware of word limit for abstract, but wonder whether you could include that information at the abstract. Alternatively, I wonder whether it would be more appropriate to cover something else at this section… after all, the study is about the feasibility of the trial, not the reliability of physical activity measures. Introduction Page 4, line 49, “comparison to waiting lists or no treatment control”: apologies, but something seems to be missing here. It is not what the limitation is. Page 4, line 52: typo “…use, [13] To our knowledge…”. Should it be full stop? Methods Page 10, line 46, “… in continuous outcome measures… for binary outcome measures”: I think this section could be revised to enhance clarity. I would suggest you to explicitly list (in brackets) what the continuous and binary outcome measures are. Results Table 4: unclear whether the number in brackets refer to 95% CI. Please revise this. Also, please define all abbreviations used in the Table. There are some sections of your results section where you are interpreting the findings. I would argue you should only report the findings, and interpret these at the discussion. Examples include: - Page 18, line 55, “However, this finding should… at baseline”: I would argue this is part of your discussion. At the results section, you should report your findings, but not interpret them. Please consider revising this. - Page 19, line 4, “Although caution is required…”: same as above. I think this should be moved to discussion. - Page 19, lines 46 to 49, “The seeming reduction… findings with caution”. Table 5: please inform the reader whether you are presenting mean values, and % of what exactly? All tables should be self-explanatory.

REVIEWER	Kay Cooper Robert Gordon University UK
REVIEW RETURNED	06-May-2017

GENERAL COMMENTS	This is an excellent manuscript reporting on the feasibility of testing an innovative internet intervention for self-management of low back pain; an area in need of high-quality studies like this. I look forward to hearing the outcome of the full-scale RCT in due course.

REVIEWER	Catherine Trask Canadian Centre for Health and Safety in Agriculture (CCHSA) University of Saskatchewan
REVIEW RETURNED	08-May-2017

GENERAL COMMENTS	This manuscript describes a feasibility trial of an internet-based back pain intervention; the protocol was published in this journal in 2015. 90 participants were randomized between 3 arms: usual care, internet-only + usual care, and internet +usual care + physiotherapy phone consult. The feasibility objectives are clearly laid out as: recruitment of at least 60 participants; 3-month attrition rate below 30%; physiotherapists completing at least 2/3 of the checklist on each call; patients access the intervention, complete session 1, and set goals; intervention is acceptable to participants. The trial was deemed both feasible on these points and trending effective; a full trial is recommended with some methodological updates. Overall this study is well-conceived, though the manuscript would benefit from some re-organization. Qualitative investigation also seems very wise given the goal of the study, and I understand that it would be too much to include a full qualitative study in this report, given the space required to present detailed methods and findings. However, it is somewhat distracting to have qualitative data listed in the study objectives and to have the interview data alluded to in the methods without any further discussion. Perhaps the fact that it was not included in this article can be stated earlier, directly after the objectives? The term 'Exploratory analyses' is used several times, and I think this merits some explanation to avoid any impression of a 'fishing expedition' that strays from the stated goals. I suspect that since this is a feasibility trial, the sample size did not support hypothesis testing, but this is worth stating explicitly in the methods analysis section. The analysis section describes linear regression models for continuous outcomes – I resume this would include the RMDQ, start back tool, tamp scale and PCS. However, data on these types of measures is rarely normally distributed, an assumption for linear regression. Were analyses carried out only on the differences, and if so, were differences normally-distributed? One of the self-reported LBP activity questions seems to be double-barreled: "Did you stop because you no longer experienced pain?" Was this question asked only of those who stopped, or was it related to a 'stop' question? It is a little difficult to interpret in the table if this is a measure of intervention success or a lack of adherence.

	Please state the clinical significance criteria of 2.0 RDMQ points in the methods section rather than citing others in your results. The results section should present only findings of the study, not interpretation, cautions, or comparison to literature. There are several cases of interpretation and discussion in the results section; I recommend moving these to the discussion. Examples: p18 lines 54-57. P19 lines 45-50 P20 lines 17-21; lines44-49 P 22, line 9: please state 'cauda equina syndrome' Economic analysis methods are not described in the methods section, and economic feasibility was not stated in the objectives. If economic analysis is important in this manuscript, please add it to the objectives and methods. Methodological description should be removed from the results section (i.e. "NHS related costs were estimated from computer records…").

REVIEWER	Lance McCracken King's College London
REVIEW RETURNED	31-May-2017

GENERAL COMMENTS	The submitted manuscript describes a feasibility trial for an internet-based treatment for low back pain set in primary care. The study is clearly described and seems to have been carefully planned and conducted. It is breaks new ground and was mainly a successful test of feasibility. It is likely to be quite interesting for researchers and clinicians otherwise interested in treatment for back pain or in internet based treatments based in primary care. I have just a few mostly minor comments. 1. It was not clear how the fidelity of the telephone support was determined. Who did this, how were they trained, and was/were their instrument/ratings reliable and valid? Similarly, how was acceptability defined? These may be in the protocol, but I think the reader would want to know these in the current paper at the point where the first encounter these issues, without having to access another article. 2. Just a thought, is it better to graph the data in Table 2? It is difficult to quickly appreciate the differences in where people discontinue from looking at the numbers alone. 3. Do the authors have any idea which of the covariates is responsible so that when it is controlled the difference at follow-up is greater for the physio supported treatment arm On the RMDQ? I'm sure the numbers are small for this type of analysis, but it is curious. 4. I may have missed it but it was not clear what measure was used to produce the "enablement scale." 5. Was it the EQ-5D-5L that was used. Is it better than the original and should it have been used? 6. There was very little description of what the treatment included. Again, it appears that further information can be gained from consulting the protocol paper. This seems inconvenient, however. For example, when it is discussed that "coping strategies" learned from SupportBack might help prevent back pain recurrence, the reader is unable to see within the submitted paper what coping strategies these might be.

REVIEWER	Anna Grimby-Ekman Health Metrics, Departmen of Medicine, Sahlgrenska Academy, Gothenburg University, Sweden
REVIEW RETURNED	31-Aug-2017

GENERAL COMMENTS	I marked that major revision is needed as I think the revision is of great impotnace for the manuscript to be consistent and structured, but most of the manuscript is really well done. Even if I marked it a major revision it is easy to do. The most important revision I think is to either extend the aim to include an aim on the clinical outcome analysis (and to explain in the manuscript why it is explorative) or to delete the results of the clinical outcome from the manuscript. Point 7 above I marked as statistical methods were appropriate and fully described. This is the case if my comment on correlation is followed. Major revision I marked that major revision is needed as I think the revision is of great importance for the manuscript to be consistent and structured, but most of the manuscript is really well done. Even if I marked it a major revision it is easy to do. The most important revision I think is to either extend the aim to include an aim on the clinical outcome analysis (and to explain in the manuscript why it is explorative) or to delete the results of the clinical outcome from the manuscript. The aim (page 2 and page 5): Clinical outcome analysis is not mentioned (the major flaw) Why “explore” the feasibility, don´t you want to determine …? Minor revisions 1. Page 2: in design and setting section the trial is described as a single center study conducted in 12 general practices? Is at one site or 12 sites the study is conducted, please explain. 2. Page 2: Why is the analysis of clinical outcomes described as explorative? Explain or consider whether it is explorative or not and how this might affect the conclusions drawn from it. 3. Page 6: Move the section “Randomization and blinding” from page 9 to page 6, and place just after section “Recruitment” 4. Page 6 (and possibly other places): Remember to refer to the protocol, ref 17, in appropriate places, for example in the randomization text, as this is very briefly described here. 5. Page 10: What kind of correlation is used? Please state. I would recommend Spearman correlation with 95% confidence interval. 6. Table 1: Take away the decimal in the percentage. You don´t have that accuracy. Also change the IQR, mentioned in first column (median days of pain …), to first (Q1) and third (Q3) quantile. The IQR is the distance between Q1 and Q3 . 7. Table 3: again the decimal in the percentages. 8. Table 4: Please add description of what numbers are presented! Means, se, 95% CI? Also make sure you have one decimal more in accuracy estimates (e.g. SE or limits og an CI) then in point estimates (e.g. mean). 9. Table 5: Again decimal in percentages. 10. Page 20: No confidence interval or even standard error! Add 95% CI to the correlation estimates.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

"The paper describes the design and results of an RCT to assess the feasibility of full trial of impact of an internet intervention to support self-management of low back pain in primary care with a primary functional outcome and 3-months.

The study is well designed and generally well presented. However, there are some issues:

Major: The authors undertake an inferential between group comparison of outcomes. This analysis is not specified in the protocol. Moreover the way this is reported in abstract (in addition to usual improved by 0.6 RMDQ points…") gives inappropriate emphasis to outcome results of what is a pilot trial that is not powered to detect a difference in outcomes. The authors should either remove this result from abstract or at least present the 95% CI around this mean estimate. The authors do draw attention to the exploratory nature of this analysis in discussion."

Response: We are grateful to the reviewer for this point. We agree that the wording emphasised outcome results and we have now removed inferential statistics from the abstract on page 2. Where we include inferential statistics in the results section we have highlighted that this reflects an addition to the protocol and is exploratory.

"Minor: The abstract needs to clarify the 3 groups who took part in the study – suggest use approach of abstract of protocol paper."

Response: We have now clarified the nature of the 3 groups early in the abstract, based on our protocol paper as suggested.

Reviewer: 2

"This study aimed to assess the feasibility of a RCT of an internet-based intervention for LBP delivered in addition to (1) usual care or (2) usual care + telephone support with a physiotherapist; and compared usual care. A total of 87 participants took part in the study, and primary outcomes were related to the feasibility of the trial (i.e. recruitment rate, adherence, and retention at follow-up). Exploratory analyses were conducted on clinical outcomes at 3 months follow-up. Results suggest that the full trial is feasible.

This is a really well-written, and nicely conducted feasibility trial. I have some minor comments for the authors, and hope these help to improve the clarity of the manuscript."

Response: We thank Reviewer 2 for their positive comments regarding our manuscript.

"Specific comments

Abstract:

Apologies, but I found the aim description a little confusing. By reading the aims of the study, I was not clear whether you designed the trial for assessing "internet + usual care" versus "usual care", and "internet + usual care + telephone support" versus "usual care". Is 'stand-alone' intervention referring to internet + usual care?

While reading the methods section of the abstract, I could not figure out what exactly the three arms were. I suggest you to revise it, to enhance clarity."

Response: We apologise for the lack of clarity in describing the arms of the trial in the abstract. Reviewer 1 makes a similar point. We have addressed Reviewer 1 and Reviewer 2's points by rewriting the objective in the abstract on page 2.

"Page 2, line 31: The term "trial time frame" is very generic, and does not give an idea of how long you recruited participants for… I think you could state what the length of period was."

Response: This is a good point. We have corrected this and now specified that participants were recruited over a 6-month period in the abstract on page 2.

"Strengths and limitations of the study

Page 3, line 16: you did not mention physical activity measurements at the abstract anywhere. I am aware of word limit for abstract, but wonder whether you could include that information at the abstract. Alternatively, I wonder whether it would be more appropriate to cover something else at this section… after all, the study is about the feasibility of the trial, not the reliability of physical activity measures."

Response: We agree with this suggestion. We have taken the point about the physical activity measure out of the strengths and limitation section, and replaced it with a broader point regarding trial methodology on page 3.

"Introduction

Page 4, line 49, "comparison to waiting lists or no treatment control": apologies, but something seems to be missing here. It is not what the limitation is."

Response: This element of a longer point in the introduction refers to the conclusions of recent systematic reviews of internet intervention for MSK pain. When internet interventions are compared to waiting list or no treatment controls, we do not know how the interventions compare to existing treatments or best practice. This reflects a limitation in study design; it makes it difficult for clinicians to know whether to recommend these interventions or not.

"Page 4, line 52: typo "…use, [13] To our knowledge…". Should it be full stop?"

Response: We have now corrected this typographical error.

"Methods

Page 10, line 46, "… in continuous outcome measures… for binary outcome measures": I think this section could be revised to enhance clarity. I would suggest you to explicitly list (in brackets) what the continuous and binary outcome measures are."

Response: We are grateful to Reviewer 2 for noting this. We have amended this section removing the mention of binary outcome measures. The majority of our measures were continuous, and where binary outcomes were explored, descriptive statistics were used (e.g. reaching MCID or not). We have added examples of continuous measures in brackets to this section on page 10. Explicitly listing all continuous measures here would duplicate the outcome and measures section found on pages 8-9.

"Results

Table 4: unclear whether the number in brackets refer to 95% CI. Please revise this. Also, please define all abbreviations used in the Table."

Response: We have now clarified that in the first two columns of Table 4, numbers refer to means (SD), and in the following 2 columns numbers refer to differences (95% CI). Additionally, we have added that the numbers for the IPAQ data refer to medians. These changes can be found on pages 16-18.

"There are some sections of your results section where you are interpreting the findings. I would argue you should only report the findings, and interpret these at the discussion. Examples include:

- Page 18, line 55, “However, this finding should… at baseline”: I would argue this is part of your discussion. At the results section, you should report your findings, but not interpret them. Please consider revising this.

- Page 19, line 4, “Although caution is required…”: same as above. I think this should be moved to discussion.
- Page 19, lines 46 to 49, “The seeming reduction… findings with caution”."

Response: These highlighted interpretive statements have now been removed from the results, and incorporated into the discussion as suggested.

"Table 5: please inform the reader whether you are presenting mean values, and % of what exactly? All tables should be self-explanatory."

Response: Apologies for the lack of clarity here, we have now amended the Table 5 label on page 19 to clarify that this table reports numbers of patients falling into each subgroup at baseline and follow-up (with the percentage).

Reviewer: 3

"This is an excellent manuscript reporting on the feasibility of testing an innovative internet intervention for self-management of low back pain; an area in need of high-quality studies like this. I look forward to hearing the outcome of the full-scale RCT in due course."

Response: We are grateful for Reviewer 3’s positive comments.

Reviewer: 4

"This manuscript describes a feasibility trial of an internet-based back pain intervention; the protocol was published in this journal in 2015. 90 participants were randomized between 3 arms: usual care, internet-only + usual care, and internet +usual care + physiotherapy phone consult. The feasibility objectives are clearly laid out as: recruitment of at least 60 participants; 3-month attrition rate below 30%; physiotherapists completing at least 2/3 of the checklist on each call; patients access the intervention, complete session 1, and set goals; intervention is acceptable to participants. The trial was deemed both feasible on these points and trending effective; a full trial is recommended with some methodological updates. Overall this study is well-conceived, though the manuscript would benefit from some re-organization.

Comment: Qualitative investigation also seems very wise given the goal of the study, and I understand that it would be too much to include a full qualitative study in this report, given the space required to present detailed methods and findings. However, it is somewhat distracting to have qualitative data listed in the study objectives and to have the interview data alluded to in the methods without any further discussion. Perhaps the fact that it was not included in this article can be stated earlier, directly after the objectives?"

Response: We thank Rewiewer 4 for this point. We now note that qualitative data will be reported elsewhere after the first mention of the qualitative aspect on page 8.

Comment: "The term ‘Exploratory analyses’ is used several times, and I think this merits some explanation to avoid any impression of a ‘fishing expedition’ that strays from the stated goals. I suspect that since this is a feasibility trial, the sample size did not support hypothesis testing, but this is worth stating explicitly in the methods analysis section."

Response: We have added text as suggested to the methods analysis section explaining what we mean by exploratory in this case. This can be found on pages 10-11.

Comment: "The analysis section describes linear regression models for continuous outcomes – I resume this would include the RMDQ, start back tool, tamp scale and PCS. However, data on these types of measures is rarely normally distributed, an assumption for linear regression. Were analyses carried out only on the differences, and if so, were differences normally-distributed?"

Response: Before using linear regression, we checked that the data met the underlying assumptions required by the model. Where this assumption was not met, for the IPAQ we used a non-parametric approach, reporting medians and the difference represents the results of a quantile regression. This has now been noted in the methods analysis section on pages 10-11.

Comment: "One of the self-reported LBP activity questions seems to be double-barreled: “Did you stop because you no longer experienced pain?” Was this question asked only of those who stopped, or was it related to a ‘stop’ question? It is a little difficult to interpret in the table if this is a measure of intervention success or a lack of adherence."

Response: We have removed this question from the table, as we agree, presented in this fashion it is hard to interpret. A question such as this may be useful in the process evaluation nested within the future full trial. The sample size in the full trial would be sufficient to examine responses to this question in the subgroup whom report low adherence; potentially providing an indication as whether low adherence was due to pain subsiding for some.

Comment: "Please state the clinical significance criteria of 2.0 RDMQ points in the methods section rather than citing others in your results."

Response: We are grateful for this point and we have now moved the explanation for this MCID to the methods analysis section, paragraph 1, page 11.

Comment: "The results section should present only findings of the study, not interpretation, cautions, or comparison to literature. There are several cases of interpretation and discussion in the results section; I recommend moving these to the discussion. Examples:
p18 lines 54-57.
P19 lines 45-50
P20 lines 17-21; lines44-49"

Following a similar point from Reviewer 2, these instances of interpretation have been removed and integrated into the discussion.

"P 22, line 9: please state ‘cauda equina syndrome’"

Response: This has now been corrected.

"Economic analysis methods are not described in the methods section, and economic feasibility was not stated in the objectives. If economic analysis is important in this manuscript, please add it to the objectives and methods. Methodological description should be removed from the results section (i.e. “NHS related costs were estimated from computer records…”)."

Response: Both of these points have now been addressed. Our HE aim has been added in paragraph 2 on page 5. We have added a HE section to the methods analysis section on page 11.

Reviewer: 5

"The submitted manuscript describes a feasibility trial for an internet-based treatment for low back pain set in primary care. The study is clearly described and seems to have been carefully planned and conducted. It is breaks new ground and was mainly a successful test of feasibility. It is likely to be quite interesting for researchers and clinicians otherwise interested in treatment for back pain or in internet based treatments based in primary care. I have just a few mostly minor comments."

Response: We grateful for this Reviewer 5’s positive comments.

"1.It was not clear how the fidelity of the telephone support was determined. Who did this, how were they trained, and was/were their instrument/ratings reliable and valid? Similarly, how was acceptability defined? These may be in the protocol, but I think the reader would want to know these in the current paper at the point where the first encounter these issues, without having to access another article."

Response: The fidelity check that was conducted has now been explained further in paragraph 2 on page 10 (whilst still aiming for brevity). All physiotherapists were provided with a checklist of topics they should aim to cover in each call. In their training, it was noted that in some calls they would not be able to cover all topics, or in some cases topics would be less appropriate for particular patients, and they should use their clinical discretion when using the guide. Thus, the fidelity check in this case was designed to broadly check that the physiotherapists were using the guide appropriately and not providing their own diagnosis and distinct suggestions. The physiotherapists experience will be explored in more depth in our associated qualitative paper.

Acceptability is now defined on page 8, where it is first mentioned with regard to the success criteria.

"2. Just a thought, is it better to graph the data in Table 2? It is difficult to quickly appreciate the differences in where people discontinue from looking at the numbers alone."

Response: We appreciate this suggestion. We tried graphing the data in a number of ways, however, we found that the graphs appeared to make the data more difficult to interpret at a glance than the original table. As such we have kept the table and tried to improve its format. The amendments can be seen on page 14.

"3. Do the authors have any idea which of the covariates is responsible so that when it is controlled the difference at follow-up is greater for the physio supported treatment arm On the RMDQ? I'm sure the numbers are small for this type of analysis, but it is curious."

Response: We agree that in a larger sample, it would be helpful to understand the extent to which any covariates are responsible for this sort of change to the difference. However, given the small numbers involved here, it is unlikely that we would be able to meaningfully undertake this sort of analysis. In a larger trial, if a similar effect is observed, we would be sure to investigate further.

"4. I may have missed it but it was not clear what measure was used to produce the "enablement scale.""

Response: We apologise, the reference for the enablement scale has now been added in paragraph 1 on page 9.

"5. Was it the EQ-5D-5L that was used. Is it better than the original and should it have been used?"

Response: On page 9 we note that it was the EQ-5D 3L that was used. We agree with this Reviewer that the EQ-5D-5L is better, and as such we will be using the EQ-5D-5L for the full trial.

"6. There was very little description of what the treatment included. Again, it appears that further information can be gained from consulting the protocol paper. This seems inconvenient, however. For example, when it is discussed that "coping strategies" learned from SupportBack might help prevent back pain recurrence, the reader is unable to see within the submitted paper what coping strategies these might be."

Response: We acknowledge Reviewer 5's point here. We have added a brief addition with which we hope to address their specific point (linking what might be learnt in SupportBack to how it might prevent recurrence) in paragraph 2 on page 7. In order to keep the manuscript near BMJ Open's recommendations on word length, we think retaining longer description of the intervention within the published protocol (also published in BMJ Open) is appropriate.

Reviewer: 6

"I marked that major revision is needed as I think the revision is of great importance for the manuscript to be consistent and structured, but most of the manuscript is really well done. Even if I marked it a major revision it is easy to do. The most important revision I think is to either extend the aim to include an aim on the clinical outcome analysis (and to explain in the manuscript why it is explorative) or to delete the results of the clinical outcome from the manuscript.
The aim (page 2 and page 5):
Clinical outcome analysis is not mentioned (the major flaw)"

Response: We are grateful to Reviewer 6 for noting this issue. The exploratory clinical outcome analysis features in the abstract in the outcome section. Following the example of recently published feasibility trials in BMJ Open, we kept our objective/aim in the abstract to a sentence. Listing all specific trial aims here would have led to us exceeding the abstract word count. We now have added our aim to explore the clinical outcome data to the expanded aims reported on page 5.

Additionally, by addressing Reviewer 1's comment and following the recently published CONSORT extension for reporting pilot and feasibility trials guidance on abstracts, we now only report the results pertaining to our primary feasibility aims our abstract.

"Why "explore" the feasibility, don´t you want to determine …?"

Response: We have now edited this so 'determine' replaces 'explore' where 'explore feasibility' appeared.

"Minor revisions

1. Page 2: in design and setting section the trial is described as a single center study conducted in 12 general practices? Is at one site or 12 sites the study is conducted, please explain."

Response: The term centre here is used to refer to single centre that recruits primary care practices (12 sites) from a single geographic area (Wessex, England). We agree this could be confusing, so have removed the single centre descriptor from the abstract on page 2.

"2. Page 2: Why is the analysis of clinical outcomes described as explorative? Explain or consider whether it is explorative or not and how this might affect the conclusions drawn from it."

Response: We appreciate this point. This has also been mentioned by other reviewers, so we now directly address what we mean by exploratory in the methods analysis section on pages 10-11.

"3. Page 6: Move the section “Randomization and blinding” from page 9 to page 6, and place just after section “Recruitment”"

Response: The current placement reflects the ordering in the CONSORT Statement, thus we would prefer it to remain. However, we would be very happy to follow to editorial guidance here.

"4. Page 6 (and possibly other places): Remember to refer to the protocol, ref 17, in appropriate places, for example in the randomization text, as this is very briefly described here."

Response: We have added a reference to the protocol to the randomisation section, and note that the method sections begins with the statement “The full details of the method and interventions can be found in the published trial protocol.[17]”

"5. Page 10: What kind of correlation is used? Please state. I would recommend Spearman correlation with 95% confidence interval."

Response: We have amended to state our use Spearman correlation and reported the 95% confidence interval, this is now mentioned in the methods analysis section on page 11.

"6. Table 1: Take away the decimal in the percentage. You don´t have that accuracy. Also change the IQR, mentioned in first column (median days of pain …), to first (Q1) and third (Q3) quantile. The IQR is the distance between Q1 and Q3 . Table 3: again the decimal in the percentages. Table 5: Again decimal in percentages"

Response: We have amended the table so that it now reads “median (Q1, Q3) “ as suggested. With respect to the accuracy with which the percentages are expressed, none of these reflect a hypothesis test and our experience is that journals prefer percentages expressed to one decimal point. However, we will be guided by the editorial policy of the journal and are happy to express percentages in all tables to the number of significant figures preferred by the editor.

"7. Table 4: Please add description of what numbers are presented! Means, se, 95% CI? Also make sure you have one decimal more in accuracy estimates (e.g. SE or limits og an CI) then in point estimates (e.g. mean)."

Response: We agree that this was not sufficiently clear and have amended Table 4 as suggested on page 16.

"8. Page 20: No confidence interval or even standard error! Add 95% CI to the correlation estimates."

Response: We are grateful to the reviewer for noting this. The Spearman correlations are now reported with 95% CI. We have also amended this section on pages 19-20 relating to correlations to ensure it is more descriptive and statements accord with the reduced sample size for these analyses.

VERSION 2 – REVIEW

REVIEWER	Daniel Ribeiro University of Otago - New Zealand
REVIEW RETURNED	25-Oct-2017

GENERAL COMMENTS	Thank you for addressing my comments during the first review round. This is a very interesting study.

REVIEWER	Anna Grimby-Ekman Health Metrics, Institution of Medicine, Gothenburg University, Sweden
REVIEW RETURNED	01-Nov-2017

GENERAL COMMENTS	A recommendation to the authors is to write a point-to-point answer to all referees in future review process. In this way you simplify the work for the reviewers and also it is respectful to answer or comment all the questions from the reviewers. It is also a help for you to see that all comments are addressed. You have improved the manuscript according to most comments, especially the major once, as I can see, but you have missed a few minor. - Most important comment: In the result section and in tables, it is not clear at all what results come from the regression analysis (linear or quantile). This must be clear both in the text in the result section and in the tables. In the Tables it should also be clearly described what confounders you adjust for. - Check all tables that the mena and the confidence intervalls don´t have same number of decimals! (old comment) The CI should have one more than the mean. See fex Table 4 (differences). - "Analysis": In the sentence "Continuous outcomes were modelled ... met the underlying assumption ..." Specify the assumptions you refer to. - Same paragraph in "Analysis": Give a reference to non-parametric quantile regression. This is not a common method, and hence not common knowledge.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2

“Thank you for addressing my comments during the first review round.
This is a very interesting study.”

We thank Reviewer 2 for their positive comments.

Reviewer: 6

“A recommendation to the authors is to write a point-to-point answer to all referees in future review process. In this way you simplify the work for the reviewers and also it is respectful to answer or comment all the questions from the reviewers. It is also a help for you to see that all comments are addressed. You have improved the manuscript according to most comments, especially the major once, as I can see, but you have missed a few minor.”

Response: We apologise to Reviewer 6 regarding the issue that some of their original comments appear to have been missed, this was not our intention. We are grateful for the minor points below, which we have now addressed.

“Most important comment: In the result section and in tables, it is not clear at all what results come from the regression analysis (linear or quantile). This must be clear both in the text in the result section and in the tables. In the Tables it should also be clearly described what confounders you adjust for.”

Response: We have added “Median (Q1, Q3)” to Table 4 on page 16 for the two outcomes that were analysed with quantile regression. There were two outcomes analysed in this way: days in pain and physical activity. We have changed the table heading so those analysed with linear models have the heading “Mean difference (95% CI)…” whilst those analysed with quantile regression have the heading “Median difference (95% CI)”. In the text for the physical activity outcome, we have also added “using quantile regression” to make this explicit. Further, we have also clarified this in title of Table 4.

In the Analysis section on pages 10-11 we have stated what all models controlled for: “each outcome at baseline, gender, age, marital status, employment status, income, ethnicity and age left education”. However, we appreciate Reviewer 6’s point, and we have added a footnote to Table 4 on page 18 to reiterate these confounders and make it clearer to a reader just looking at the tables.

“Check all tables that the mena and the confidence intervalls don´t have same number of decimals! (old comment) The CI should have one more than the mean. See fex Table 4 (differences).”

Response: We have amended the tables accordingly.

“"Analysis": In the sentence "Continuous outcomes were modelled ... met the underlying assumption ..." Specify the assumptions you refer to.”

Response: We have amended the text on pages 10-11 to read: “if they met the underlying assumptions that the outcome measure and the residuals were normally distributed".

“Same paragraph in "Analysis": Give a reference to non-parametric quantile regression. This is not a common method, and hence not common knowledge”

Response: We have added the reference to Hao and Naiman’s 2007 book 'Quantile regression' on page 11.