Author’s response to reviews

Title: The association between believing staying active is beneficial and achieving a clinically relevant functional improvement after 52 weeks: a prospective cohort study of patients with chronic low back pain in secondary care

Authors:

Allan Riis (ariis@dcm.aau.dk; a.riis@rn.dk)

Emma Karran (Emma.Karran@mymail.unisa.edu.au)

Janus Thomsen (jlt@dcm.aau.dk)

Anette Jørgensen (anetjoer@rm.dk)

Søren Holst (Soeren.Holst@midt.rm.dk)

Nanna Rolving (Nanna.Rolving@rm.dk)

Version: 1 Date: 20 Dec 2019

Author's response to reviews:

Dear Dr. Kristin Archer

Thank you for considering manuscript BMSD-D-19-01422: ‘Does disease perception influence functional outcome among patients with low back pain? A prospective cohort study with 52 weeks follow-up’ for publication in BMC Musculoskeletal Disorders. We have responded to the comments below and uploaded a manuscript with changes in red and a clean version of the manuscript.

**********

Thank you for submitting your manuscript for consideration in BMC Musculoskeletal Disorders. We appreciate the effort that went into your manuscript. Both the reviewers and I agree that this paper addresses an important topic and has the potential to contribute to knowledge in the field. However, there are many important issues that need to be addressed and a major revision is necessary. The reviewers have provided guidance for your revision; I have some additional suggestions that should be considered. I am summarizing our major concerns below.

1. I agree with both reviewers that the central concept for this study needs to be clarified throughout the manuscript. Based on the introduction and hypothesis, the study is focusing on patient beliefs about staying active and the relationship between these beliefs and functional outcome (i.e., RMDQ) at 52 weeks after referral to an outpatient spine clinic. However, the methodological approach does not fully support the intent and primary aim of the study.
Additional clarification is needed on the selection of independent variables, covariates, outcomes, and the modeling approach for all analyses.

a. Several secondary explanatory variables are included in this study; however, it is unclear why these were selected. The questions about finding the cause of the pain, importance of x-rays and MRI scans, practitioner advice, physiotherapist advice, pain duration, pain rating, and STarT screening tool do not appear to directly relate to patient beliefs about activity. Please provide a rationale for including these variables in the study and clarify in the methods section how all of these variables are being used in the analyses.

Response 1a: The overall purpose and the rationale for including additional explanatory variables is added in the background section page 5, line 86: ‘Consequently, the overall purpose is to study if guideline concordant information and advice provided by health care professionals are associated to better patient outcomes. Guideline guidelines include advice to stay, do not recommend x-rays and MR scans to rule in, recommend informing patients that finding the cause is often not relevant for recovery, and recommend to avoid referrals to specialist care for patients with acute or mild pain [6].’

b. In particular, I agree with Reviewer 1 that a rationale for the inclusion of the STarT screening tool as a main independent variable needs to be provided. The introduction is focused mainly on beliefs about activity and does not provide important background information on the STarT tool. Since the STarT tool is a major topic of the discussion, the study aim and hypothesis pertaining to the STarT tool should be provided. A detailed description of the STarT tools is also needed in the methods section.

Response 1b: A rationale for including STarT Back Tool is now provided in the background section page 4, line 79: ‘Patients with multiple psychological barriers for recovery who are classified as ‘high risk’ according to The STarT Back screening tool are expected to have a poorer outcome than patients not in the high risk group [14-15]. More complex care may be indicated for this high risk group – including specialist consultation – such that their identification was considered relevant to this study’

Furthermore, a detailed description of the STarT Back Tool is included in the methods section on page 8, line 168: ‘The STarT Back Tool is a patient reported questionnaire, which allocates respondents with LBP to low risk, medium risk, or high risk categories based on nine items [23]. Patients in the high-risk group are characterised by their holding a combination multiple psychological barriers (beliefs) to recovery (items 5-9), e.g. being fear avoidant, having worrying thoughts about recovery, having low mood, and/or catastrophizing [23].’

c. A multitude of important baseline characteristics are collected and displayed in Table 1. However, only age, education, and gender are controlled for in the analyses. The limited adjustment for confounding variables is a major limitation of the study. It is unclear why additional confounders, especially pain duration and pain intensity as well as smoking, employment, and comorbidities, were not included in the models.
Response 1c: We did collect multitude of baseline characteristics. Our primary concern for not including all variables in one analysis was to overcorrect estimates by adjusting for intermediate variables and thereby introducing bias of estimates. Since, some variables can be considered steps in a casual pathway: Receiving advice to stay active, believing staying active well helps recovery, avoiding sick leave, obtaining a better RMDQ score. On page 14, line 296 we have added: ‘In the statistical analysis we described the planned collection of data for patients’ health status (e.g. functional level and pain), behaviours (e.g. sick leave and smoking), beliefs (e.g. about staying active and importance of scans), and treatment received (e.g. receiving advice to stay active) [28]. We made an a priori decision to adjust for age, gender, and level of education [28] to avoid the risk of overcorrection of confounders by adjusting for intermediate variables [33]. A post hoc analysis including baseline RMDQ, pain duration, pain intensity, smoking, employment, and comorbidities, estimates did not significantly change in size or direction of estimates.’

d. Please provide a rationale for including secondary outcomes of pain and physical activity. The authors do not have an aim or hypothesis for these outcomes, statistical analysis for these outcomes are not provided, and these outcomes (Table 3) are not discussed in the results or discussion section. Please consider removing these outcomes if not relevant to study aims and data will not be provided or discussed in the paper.

Response 1d: The secondary outcomes (Pain and physical activity) are removed from the methods section and, consequently, Table 3 is removed from the manuscript.

Comment 2. I agree with Reviewer 1 that additional clarification is needed for using a threshold of 30% improvement for the primary outcome (RMDQ)

Response 2: We dichotomized the RMDQ score at the limit of a clinical relevant improvement. The text on page 9, line 180 now reads: ‘A threshold of 30% improvement between baseline and 52 weeks was considered a clinically relevant improvement [26]. This was based on study from the UK, where a 30% improvement in RMDQ between baseline and follow-up defined the clinically relevant improvement in function [27].’

Comment 3. Please clarify all the analyses that are being conducted for this study in the statistical analysis section. Based on the secondary explanatory variables and Table 2, there appears to be 8 separate regression analyses that are being conducted for the RMDQ outcome. The rationale for all eight of these analyses is unclear since the study has 1 aim and hypothesis that focuses on beliefs about activity. In addition, it is unclear why there are so many separate analyses when these variables can all be included in the same analysis. Finally, there are some interesting data presented in Tables 1 and 2; however, these data are not fully explored in the results and discussion section.

Response 3: We have now included the additional aims in the aim section of the manuscript and the rationale for including several aims/analyses was provided above (Comment 1a). A concern in this study was to overcorrect for confounders, that is the reason for only adjusting for age,
gender, and educational level (As stated in the SAP). Following your comment and the previous changes, the beginning of the statistical analysis section is now on page 9, line 189 and has been changed to ‘Responses (0-5 or 6-10) to the question ‘An increase in pain is an indication that I should stop what I’m doing until the pain decreases’ was considered the primary explanatory variable. Responses to ‘I think that finding the cause of pain is important for my recovery’, ‘I think x-rays and MR scans are an important part of my recovery’, 'Have you been advised by your general practitioner to stay active despite your back pain?’, Have you been advised by a physiotherapist or chiropractor to stay active despite your back pain?’, pain duration (≥ 12 weeks), numerical pain rating (0-10), and the STarT Back Tool (High risk group) were all considered secondary explanatory variables. The outcome (RMDQ) was adjusted for possible confounders (age, gender, and educational level) in a logistic regression analysis and analysed unadjusted in logistic regression models estimating the odds of clinical relevant improvement (30%) in the RMDQ score (Figure 1). All explanatory variables are reported as adjusted and unadjusted odds ratios, supported by numbers of patients with clinically relevant improvements.’

Furthermore, all results are now included in the result section of the manuscript. On page 12, line 255, we added: ‘Pain duration was statistically significantly associated with a clinically relevant improvement in the RMDQ (P<0.001). Numerical pain rating (P=0.069) and advice to stay active from a GP (P=0.218) or a physiotherapist or chiropractor (P=0.933) were not found to be associated to an improvement in RMDQ.’

Comment 4. Please clarify why the power calculation section includes information on outcome differences between chronic and pain free patients when this does not appear to be an aim of the study. The power calculation section does not appear to address the main aim of the study and the modeling approach selected.

Response 4: The power calculation is rewritten on page 10, line 213: ‘In a study from New Zealand, 80% of patients with LBP agreed with ‘If you have back pain you should try to stay active’ and respondents holding more positive views about activity if they had consulted a health professional [29]. In our sample, all patients were referred from primary care and had consulted at least one health professional. Consequently, 33% was considered a realistic estimate of the proportion of patients agreeing with: ‘An increase in pain is an indication that I should stop what I’m doing until the pain decreases’ (reply 6-10). 67% were predicted to disagree (reply 0-5) with ‘An increase in pain is an indication that I should stop what I’m doing until the pain decreases’. Among patients replying (6-10), 50% were expected to have a clinically-relevant improvement in their RMDQ score [27]. Among patients replying [0-5], 70% were expected to receive a clinically relevant improvement in their RMDQ score. With full follow-up, group sizes (33/67), alpha 0.01, and a power of 0.9; 423 patients were needed in the analysis. To account for different group sizes and loss to follow-up 800 patients needed to be recruited.’

Comment 5. In the first paragraph of the discussion, the authors state that patients with shorter pain duration have better outcomes. It is unclear why the authors are including this as a main finding since these data were not provided in the results section and this was not an aim of the
study. Pain duration has the potential to be a confounding variable and should be discussed in relation to the association between activity beliefs and the functional outcome of RMDQ.

Response 5: Following the previous comments and responses, pain duration is now included as an additional aim and is presented in the result section of the manuscript. Furthermore, page 14, line 300 now reads ‘A post hoc analysis including age, gender, educational level, baseline RMDQ, pain duration, pain intensity, smoking, employment, and comorbidities, estimates did not significantly change the size or direction of estimates.’

Comment 6. The discussion section (page 13, lines 37-40) includes information about medium or low risk patients and their relevant functional improvement; however, these data are not provided in the Tables or results section.

Response 6: We have now avoided discussing the STarT Back Tool low risk group and the high risk group in the discussion. The sentence is rephrased and a reference to Table 2 is inserted. In Table 2 the odds ratio for the association between being in the STarT Back Tool high risk group and RMDQ is presented.

The text in the discussion section on page 13, line 264 now reads: ‘Holding multiple erroneous beliefs, (presence of fear-avoidance, catastrophic thoughts, and depressive thoughts) and thereby being in the STarT Back Tool high-risk group [14] was associated with not achieving a clinically relevant improvement in function at 52 weeks (Table 2)’

Comment 7. Please consider revising the discussion and conclusion section to include all relevant data that are presented descriptively and as multivariable regression results.

Response 7: Following all comments, the discussion section and the conclusion section in the manuscript are revised.

****************

Jason Beneciuk (Reviewer 1):

Thank you for the opportunity to review this manuscript involving relationships between patient beliefs about pain as an indicator for continued activity and disability outcomes at 52 weeks. There are several strengths to this study including large sample size, adequate follow-up rate, and use of STROBE guidelines. Although this topic is important, there were several concerns described in detail below.

Major comments:

Comment 8: It is not clear what the primary implications are related to the main take home message (i.e., pain as a warning sign to stop being active; was not associated with less favorable
outcomes). Realizing this was an observational study, I was still looking for some speculation as to how these results may influence clinical practice - or is there something else that needs to be emphasized?

Response 8: New text is added to the ‘Meaning of the study: possible mechanisms and implications for clinicians or policymakers’ section, page 16, line 342: ‘Advice to stay active is a cornerstone of clinical guidelines [3-6] yet the effects of such advice is unclear. This study indicates that a single belief about physical activity and recovery cannot predict patients’ prognosis in terms of functional ability (RMDQ) among patients in general practice referred to hospital care. Instead, having a range of unhelpful beliefs, as measured with the STarT Back Tool, seemed to be more closely associated with functional capacity after 52 weeks. This study emphasises the need to address multiple psychological beliefs including patients’ fear-avoidant beliefs to obtain better treatment effects in clinical practice, and suggests that simply delivering the advice to stay active might not change functional outcome.’

Comment 9: Related to prior comment, the authors acknowledge the importance of patient beliefs about clinical outcomes is not well understood (Background, line 54); however not completely sure what this study adds to the literature.

Response 9: This comment is closely related to comment 8, page 16, line 347 now reads ‘This study emphasises the need to address multiple psychological beliefs including patients’ fear-avoidant beliefs to obtain better treatment effects in clinical practice, and suggests that simply delivering the advice to stay active might not change functional outcome.’

Comment 10: Time line between initial appointment and 52 weeks is extensive and does not account for other potentially confounding factors (e.g., recurrence, new episodes).

Response 10: We agree with not including intermediate follow-up is a limitation of the study. Page 14, line 288: ‘Intermediate outcome measurements between baseline and 52 weeks was not collected. This is a limitation, since intermediate measurement may have identified potential confounding from recurrence and new episodes.’

Comment 11: Baseline questionnaire to assess "agreement" with specific statements or questions is unclear (e.g., not relevant to items 4-8). Related there is not rationale for why agreement outcomes were dichotomized.

Response 11: Argument for including items 4-8 is provided (comment 1a) and we do find it interesting to distinguish between the belief in the need to stay active and the experience of being recommended to stay active by a healthcare professional. Consequently, we prefer to keep items 4-8 in the manuscript. However, we will remove items 4-8 on the editors’ request.
Comment 12: Rationale for using the NPRS as an explanatory variable and outcome is appropriate if controlling for baseline scores, however not clear why the same strategy was not used for baseline RMDQ scores.

Response 12: The argument for keeping pain as an explanatory variable is provided in comment 1a and pain is removed as outcome (comment 1d). Adjustment for baseline RMDQ was not planned in the protocol. However, we do believe this can provide important information. Page 14, line 300: ‘A post hoc analysis including age, gender, educational level, baseline RMDQ, pain duration, pain intensity, smoking, employment, and comorbidities, estimates did not significantly change the size or direction of estimates.’

Comment 13: The fact that 90% had chronic low back pain needs to be reflected in title - and may also contribute to patient beliefs. This topic is not addressed in Discussion.

Response 13: The title is changed to ‘The association between believing staying active is beneficial and achieving a clinical relevant functional improvement after 52 weeks: a prospective cohort study of patients with chronic low back pain in secondary care’

The topic is now addressed in the discussion section, page 13, line 269: ‘The majority of patients had pain for 3 months or longer. The minority of patients with shorter pain durations reported better outcomes, which is in line with previous findings [30]. Furthermore (as mentioned above), page 14, line 300 now reads ‘A post hoc analysis including age, gender, educational level, baseline RMDQ, pain duration, pain intensity, smoking, employment, and comorbidities, estimates did not significantly change the size or direction of estimates.’

Comment 14: Regarding the delivery of information delivered to patients: 1) how was 'effective' delivery assessed? And 2) how was the information delivered? (Background, line 38)...Related to the delivery of information (if it was not assessed, which my initial impression suggests it was not), the authors need to reconsider suggestions provided in "meaning of the study" (i.e., page 14, line 9) "This suggests that simply informing patients…"

Response 14: Thank you for pointing to this. On page 16, line 335, the text is changed to ‘Possibly just believing staying active is helpful for recovery is not enough to change outcomes and a more complex psycho-behavioural approach may be required to improve patient outcomes.’

Minor comments:

Comment 15: It was not clear what the rationale was for dichotomizing the RMD (primary outcome).

Response 15: We dichotomized the RMDQ score at the limit of a clinically relevant improvement. The text on page 9, line 180 now reads: ‘A threshold of 30% improvement
between baseline and 52 weeks was considered a clinically relevant improvement [26]. This was based on a study from the UK, where a 30% improvement in RMDQ between baseline and follow-up defined the clinically relevant improvement in function [27].

Comment 16: Methods to assess for face validity (data collection, line 52) with senior author present is a bit concerning and modifications were not discussed.

Response 16: Question wording and measurement properties have previously been tested, page 13, line 275: ‘This Danish version of the Örebro has been validated as a complete questionnaire [21], it showed acceptable measurement properties in terms of test-retest reliability and absolute reliability [21].’

However, choosing one item from the questionnaire context might change respondents’ ability to understand and respond to the item. Consequently, face validity was tested on 10 patients, page 13, line 277:

‘The applied single item was pilot-tested before the trial for face validity by NR in 10 patients, who understood the question and were able to reply.’

Comment 17: STarT Back Tool ‘high risk’ group findings are not surprising and not sure how including the SBT contributes to this study.

Response 17: We agree that this is not surprising, however, it is important for clinical management. Page 16, line 342 now reads ‘Advice to stay active is a cornerstone of clinical guidelines [3-6] yet the effects of such advice is unclear. This study indicates that a single belief about physical activity and recovery cannot predict patients’ prognosis in terms of functional ability (RMDQ) among patients in general practice referred to hospital care. Instead, having a range of unhelpful beliefs, as measured with the STarT Back Tool, seemed to be more closely associated with functional capacity after 52 weeks. This study emphasises the need to address multiple psychological beliefs including patients’ fear-avoidant beliefs to obtain better treatment effects in clinical practice, and suggests that simply delivering the advice to stay active might not change functional outcome’

**

Mari Lundberg (Reviewer 2):

General comments

I like the idea, but the paper suffers from methodological inconsistencies that needs to be clarified and elaborated on. My major concern is the use of one single question as the primary outcome without having strong enough support for its validity.

My major concerns are the following.
The rationale for the study

Conceptual definitions vs operational definitions

Comment 18: Central concepts. The central concepts need to be clearly defined. As for now it reads "disease perception" in the title, but in the purpose the primary question is about "believing that staying active is beneficial despite having LBP", and in the methods it is measured by question 9 in the Örebro musculoskeletal short form. So what is it that you want to measure? I do not consider it correct that the question you have chosen reflects "disease perception". Please clarify.

Response 18: Thank you for pointing to this. The title is changed to ‘The association between believing staying active is beneficial and achieving a clinical relevant functional improvement after 52 weeks: a prospective cohort study of patients with chronic low back pain in secondary care’

The aim section of the manuscript on page 5, line 95 now reads ‘The main aim of this study was to investigate if believing staying active despite having pain is associated to better functional outcomes among patients referred from general practice to an outpatient spine clinic in a secondary care setting in Denmark. We hypothesised that patients who believe that staying active will help them recover will have higher odds of a 30%-improvement in the Roland Morris Disability Questionnaire (RMDQ) score after 52 weeks compared to patients who do not believe that staying active will help them recover. Additional aims were to study the association between having received advice to stay active by a health professional, pain intensity, pain duration, and the STarT Back Tool and recovery in the RMDQ score after 52 weeks.’

And the aim in the abstract is changed to ‘The main aim was to investigate if believing staying active despite having pain is associated to better functional outcomes after 52 weeks.’

Comment 19: Valid measures. In relation to the conceptual definition, you also need to clarify the validity of that one question in the Örebro questionnaire. That is, is that only question valid to measure the concept you want to measure? In the discussion you write to have tested with a face validity procedure, and that it did not gave any more information to apply cut-off points. How was that tested, and based on what arguments did you come to that conclusion?

Response 19: Item 9 in the short form Örebro questionnaire was used as an explanatory variable. The Danish version has previously been validated including tested for face validity. The text in the ‘Strengths and weaknesses of the study’ section now reads on page 13, line 274: ‘We applied a one-item global rating (0-10) question to measure the patients’ views on the importance of staying active (part of the Örebro questionnaire [19]). This Danish version of the Örebro has been validated as a complete questionnaire [21], it showed acceptable measurement properties in terms of test-retest reliability and absolute reliability [21]. The applied single item was pilot-tested before the trial for face validity by NR in 10 patients, who understood the question and were able to reply’.

Data management and statistical management
Comment 20: Dichotomizing. You have chosen to dichotomize the rating on the Örebro scale according to the following; Scores of 0-5 were coded as 'disagree' and scores of 6-10 were coded as 'agree'. In for example the use of an NRS scale for pain, one chooses to divide the responses according to the picture presented below. What are your arguments for choosing to do so, and what is the risk with dichotomizing?

Response 20: We did not expect the score to be normally distributed and we wanted to adjust for age, gender, and educational level, therefore, logistic regression was considered a feasible method. In the ‘Strengths and weaknesses of the study’ section on page 13, line 279 we have added the sentences: ‘Applying a cut-off between 5 and 6 was considered a practical solution and is supported by findings from Norway, where patients having a high psychosocial sub-score according to SBT expressed a median fear of working 10 (iqr, 6-10) [31]. Visual inspection showed that the distribution was not normally distributed and did not reveal the distribution of three or more subgroups. Applying other cut-off points did not considerably change the estimates. Even though, dichotomizing in general can lead to loss of information and thereby loss of power [32], dichotomizing of the primary explanatory was found optimal in our analysis.’

Comment 21: Logical reasoning. How did you come to the conclusion that 33% would respond "If pain is increasing they should stop with their physical activities", that is what are your arguments for that number?

Response 21: We wanted to allow for uneven group sizes in the power calculation. The text in the power calculation section on page 10, line 213 now reads: ‘In a study from New Zealand, 80% of patients with LBP agreed with ‘If you have back pain you should try to stay active’ and respondents holding more positive views about activity if they had consulted a health professional [29]. In our sample, all patients were referred from primary care and had consulted at least one health professional. Consequently, 33% was considered a realistic estimate of the proportion of patients agreeing with: ‘An increase in pain is an indication that I should stop what I’m doing until the pain decreases’ (reply 6-10). 67% were predicted to disagree (reply 0-5) with ‘An increase in pain is an indication that I should stop what I’m doing until the pain decreases’. Among patients replying (6-10), 50% were expected to have a clinically-relevant improvement in their RMDQ score [27]. Among patients replying [0-5], 70% were expected to receive a clinically relevant improvement in their RMDQ score. With full follow-up, group sizes (33/67), alpha 0.01, and a power of 0.9; 423 patients were needed in the analysis. To account for different group sizes and loss to follow-up 800 patients needed to be recruited.’

Comment 22: Missing. What are the consequences of your high rate of external missing, that is the number who did not respond to the questionnaire at all? That needs to be responded to and also accounted for in your discussion and conclusion.

Response 22: We have included a declaration of this limitation in the discussion section on page 14, line 290: ‘Of the 1,789 patients invited, 828 (46%) accepted to participate. We did not collect consent to report on baseline characteristics among non-participants, this may limit the generalisability of the findings and is a weakness of the study.’
Comment 23: Language. Your language needs to be revised in accordance with scientific writing. In the conclusion you write "patients classified as high risk according to the STarT Back Tool appeared to indeed be at higher risk of a poor outcome after 52 weeks". Indeed needs to be taken away.

Response 23: Thanks for pointing to this, the conclusion is rewritten an “indeed” is removed. The manuscript has been proof-read by a proof reading company.

Comment 24: Tables and Figures. All tables and figures need to be revised both in formatting and content. For example, the columns should have the same width and the text should be consistently presented (either indented to the right, left or centred). The abbreviations of sd etc should preferable be moved to the presentation of the variable and not in the result itself. The Tables and Figures have duplicate headings, which needs to be adjusted. These are just some examples. Read the instructions thoroughly.

Response 24: Figure and Tables have been formatted.

On behalf of all authors,

Allan Riis