## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Does the performance of five back-associated exercises relate to the presence of low back pain? A cross-sectional observational investigation in regional Australian council workers |
| **AUTHORS** | Gabel, Charles Philip; Mokhtarinia, Hamid; Hoffman, Jonathan; Osborne, Jason; Laakso, Liisa; Melloh, M |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Prawit Janwantanakul<br>Department of Physical Therapy, Faculty of Allied Heath Sciences, Chulalongkorn University, Bangkok, Thailand |
| **REVIEW RETURNED** | 27-Jan-2018 |

| | |
|---|---|
| **GENERAL COMMENTS** | General comments<br>==========<br>This cross-sectional study aimed to evaluate the relationship between the ability/inability to perform five physical test-exercises and the presence and/or absence of LBP. The topic is of interest and the paper is well-written. My concern is mainly about insufficient details in the methodology and result sections of the paper. As a result, it is difficult to comprehend the findings and to consider its scientific strength.<br><br>Specific comments<br>==========<br>1. Abstract<br>•Would it be correct to generalize the findings to a general working population, despite data were collected from a population of employees with the Sunshine Coast Regional Council in Queensland, Australia?<br>•It is unclear about the meaning of “reductions in LBP”. Incidence or severity of LBP?<br>•This is prospective cross-sectional, exploratory, observational investigation. What is ‘prospective cross-sectional’?<br><br>2. INTRODUCTION<br>•“…Cross-sectional analysis of these working groups (a population of employees with the Sunshine Coast Regional Council in Queensland, Australia) can be representative of general working populations…” Is it a valid statement?<br>•“As a consequence of the knowledge gap in the research of these modifiable factors, there is a need for an observational study in a representative working population to ascertain and analyze the relationship between the reported presence of LBP symptoms and the individuals physical functional movement capabilities”. It is not so clear to me about the difference(s) between this study and previous studies regarding predictive value of physical functional capabilities |

| | |
|---|---|
| | on the onset or recurrence of LBP.<br>•"Analysis of the findings may indicate what movements, or lack of movements, might be associated with the presence and/or absence of LBP. The outcomes may contribute to both the understanding of the relevance of functional movement and exercises as well as provide direction for future prospective studies..." My understanding is that numerous structures in the low back and nearby regions may be the sources of LBP. Thus, there is no 'one size fits all' intervention to prevent and treat LBP. Therefore, implementing the same preventive regime for all to prevent and treat LBP would be irrational.<br><br>3. METHODS<br>•There is no information about the inclusion and exclusion criteria of participants into the study.<br>•The rationale for including these exercises and the objective(s) of each exercise should be provided. Details about the criteria to identify those can and cannot perform each of five exercises should also be included.<br>•Each participant also completed a questionnaire: 'How often do you have low back pain?'. What was the timeframe for the question, e.g. 1 year or lifetime?<br>•"This 'LBP' group was, subsequently, further dichotomized into 'Some' and 'Most' to sub-categorize the severity of LBP being present in their lives 'sometimes' or 'most of/all the time' (Some/Most)…" I am a bit confused with this statement. What were the response options correlated with the number of exercises the individual was able to perform successfully?<br>•There is no information about the test-retest reliability of performing the exercises. Poor test-retest reliability would compromise the internal validity of the study.<br>•Details about statistical analysis are essentially required.<br><br>4. RESULTS<br>•Demographic information of participants is essentially required.<br>•How about the relationship between each exercise and the presence and/or absence of LBP? This information, similar to Table 1, would be helpful to understand the results in Table 4 and 5.<br>•"…less than 3% of all participants were able to complete one or no exercises…" This sentence is quite difficult to comprehend.<br>•It is difficult to understand the findings because no information about statistical analysis is provided. For example, no information about confounders included in the logistic regression analysis and how they were derived is available. Why was the gender effect tested only?<br><br>5. DISCUSSION<br>•"These exercises could be used clinically to diagnose the potential severity of LBP, and perhaps severity of impairment" This sentence was not supported by the findings of this study, i.e. no data collection regarding pain severity or impairment.<br>•"Physical functional tests, especially those emphasized in this study, are directed primarily toward the abdominal and lumbo-pelvic muscles and their coordinated activity" LBP in different occupations is unlikely to originate from identical causes because patients are exposed to different risk factors. In this case, is it possible that lumbar stability may not be the problem for everyone? Some discussion to acknowledge this fact would be useful.<br>•I am not sure that the authors provided possible explanations or hypotheses for why the number of exercises completed might |

| | |
|---|---|
| | predict the presence and/or absence of LBP in the sample of population.<br>•"Strengths of this study include the prospective nature…" Is this study cross-sectional in design?<br>•"…with diverse age groups and occupations…" and "…degree of homogeneity in the sample…" seems to contradict each other.<br>•One of the limitations of the study is its cross sectional design which does not allow for the causal relationship to be investigated. Also, a subjective diagnosis of LBP is another study weakness.<br><br>6. Others<br>•Below Table 1 should add the meaning of 1, 2, and 3.<br>•Would it be easier to comprehend Table 3 and 5 by modifying the current (statistical) format to simpler format? |

| | |
|---|---|
| **REVIEWER** | Achim Elfering<br>University of Bern, Switzerland |
| **REVIEW RETURNED** | 05-Feb-2018 |

| | |
|---|---|
| **GENERAL COMMENTS** | Bmjopen-2017-020946<br>Theory<br>The ms reports an inverse association between the capacity to perform 5 exercises and self-reported frequency of back pain. Data are cross-sectional. The association is worth reporting – however the author(s) should be aware that causality can not be inferred. Causality would need a temporal sequence of antecedent and consequence and causality would need to rule out alternative explanations. The author(s) aknowledge this shortcoming but sometimes there is also some « causal » writing , e.g., in admitting a „prospective nature" of the study on page 7 line 58. Indeed there is evidence that back pain reduces capacity and low capacity increases risk for back pain and third variables (e.g., fittness, health status, comorbidities) may affect both capacity and back pain.<br>The author(s) should consider such models in more depth in the introduction and in the discussion.<br>It would be interesting to control for potential third variables (e.g., fittness, etc. ) in analyses. It would also be interesting to see whether exercises are uniquely related to back pain over and above self-report screening information from questionnaires. If author(s) have the information they should add analyses to the ms.<br><br>Methods<br>Please report the participation rate: How many participants of educational classes agreed to do the exercises?<br>Please report the context of exercise. What was the instruction ?<br>Which criteria were chosen to define an exercise to be done or failed ? Please report reliability estimates on these criteria.<br>Were the exercise done alone or in presence of other class members ?<br>Were the exercise done before or after filling out the questionnaire on low back pain ?<br>Whyh did the author(s) cite a quationnaire on attention-deficit/hyperactivity disorder (ref 55) for their own question on back pain<br>Please report the sensitivity, specificity and predictive value of exercises. |

**VERSION 1 – AUTHOR RESPONSE**

Responses to Reviewer: 1

1. General comments
==========
This cross-sectional study aimed to evaluate the relationship between the ability/inability to perform five physical test-exercises and the presence and/or absence of LBP. The topic is of interest and the paper is well-written. My concern is mainly about insufficient details in the methodology and result sections of the paper. As a result, it is difficult to comprehend the findings and to consider its scientific strength.

We thank the reviewer and have accordingly modified each section in direct response to the comments provided.
We feel we have addressed the concerns regarding detail related particularly to the Methods and Results and we believe that the changes as suggested clarify matters to ensure the scientific strength is more readily comprehensible.

Specific comments
==========
1. Abstract
• Would it be correct to generalize the findings to a general working population, despite data were collected from a population of employees with the Sunshine Coast Regional Council in Queensland, Australia?
- Yes.
- We feel generalizability can be implied as the cross section of occupations, genders and age groups are respectively extensive and diverse with the population at n=422 of an initial n=539.
- We have provided the demographics in the new 'Table 1' with age range and variation, gender and occupation type - detailing 21 separate occupations plus an 'Other' category.
o There is a weighting towards manual and outdoor workers over sedentary or office workers. However, it is noted that female dominated child care and community services, are manual and non-sedentary.
o Consequently we feel that the data supports an implication of generalization.
• It is unclear about the meaning of "reductions in LBP". Incidence or severity of LBP?
Clarified to incidence
• This is prospective cross-sectional, exploratory, observational investigation. What is 'prospective cross-sectional'?
Changed to remove prospective - cross-sectional

2. INTRODUCTION
• "…Cross-sectional analysis of these working groups (a population of employees with the Sunshine Coast Regional Council in Queensland, Australia) can be representative of general working populations…" Is it a valid statement?-
- Yes. We feel generalizability can be implied as the cross section of occupations, genders and age groups are respectively extensive and diverse with the population at n=422 of an initial n=539.

• "As a consequence of the knowledge gap in the research of these modifiable factors, there is a need for an observational study in a representative working population to ascertain and analyze the relationship between the reported presence of LBP symptoms and the individuals physical functional movement capabilities". It is not so clear to me about the difference(s) between this study and previous studies regarding predictive value of physical functional capabilities on the onset or recurrence of LBP.

We thank the reviewer for these insightful comments and support our previous position of the ability to

imply representation of general working populations. We have modified the text to reflect this matter and the reviewer's comments with the understanding of generalizability within the frame work below; we have also modified the section in the introduction and discussion pertaining to previous studies on physical functional capabilities to expand and clarify:
Representative Working populations and Generalizability:
- The provision of the occupational representation in Table 1 implies support for generalization as the occupations include a significant cross-section of occupations, weighted toward those with higher risk in the manual heavy and repeated loads, e.g. road repair workers, parks maintenance, child care, community service construction, and infrastructure – but also considering the cumulative load and risks of individuals who are sedentary in office based organizational, technology, and management positions.
We have cited the references of:
- Riihimäki (1989) who considered generalizability in their 'Low-back pain and occupation. A cross-sectional questionnaire study of men in machine -operating, dynamic physical work, and sedentary work' in Spine where the variation in occupational loads was made through a diverse population by using machine operating dynamic physical work and sedentary work - similar to the approach of our study.

- More specifically Laoa (2016) considered a representative population in their study in the Safety Science Journal on 'Working smart: An exploration of council workers' experiences and perceptions of heat in Adelaide, South Australia' and, consequently, specifically selected council workers as the representative population.

- Trask (2016) highlights the importance of 'Building a Foundation for Epidemiological Studies' in this planning article in Journal of Medical Internet Research –Protocols (JMIR-Res Protoc) and discussed the need for studies to have varied geographical settings and occupational determinants in order to facilitate future research evaluating the effectiveness of safety measures to imply generalizability and general population representation.

- Coenen (2014) in Occup Environ Med discusses the relevance of 'Cumulative mechanical low-back load at work is a determinant of low-back pain' due potentially to microdamage accumulation or fatigue especially in 19 groups of highly exposed occupational groups handling heavy loads and in awkward body postures. This call for occupational diversity in studies is reflected in our study as a way to work towards generalizability from the sample population investigated and that it can be representative of the general working population.
To better simply rather than verify that there is generalizability and that a representative population was used we have modified the text as follows:
Existing research has a knowledge-gap for modifiable factors demonstrating a need for observational studies in representative working populations. Addressing this gap will assist in identifying the relationship between LBP symptoms and individual physical functional movement capabilities. A representative group, with strong indicators of generalizability, is council workers. The group includes diversity of gender, age and occupations with variance in manual and sustained loads{Riihimäki, 1989 #3498} and stationary and sedentary postures{Laoa, 2016 #3491}. Cross-sectional analysis of these groups is a starting point in implied generalization and provides insight into the capacities and abilities that may lead to the presence or risk of LBP{Trask, 2016 #3489;Coenen, 2014 #3490}.

Previous studies on physical functional capabilities
We modified the introduction /discussion on previous studies on physical functional capabilities to expand and clarify:
Introduction
… In contrast, modifiable factors{Beneciuk, 2013 #2543;Mitchell, 2010 #1891} including movement patterns{Hoffman, 2016 #3257;O'Sullivan, 2005 #3256}, physiological loads{Hoffman, 2013 #2560},

and exercise capacity{Micheo, 2012 #3223;Grönblad, 1994 #3497} receive limited attention yet they significantly influence LBP morbidity and symptomology{Deyo, 2001 #3492;Balagué, 2012 #3618}, being recognized as potentially able to prevent LBP{Hancock, 2015 #3119}.
….
The conundrum remains that LBP reduces functional capacity, fitness and general health status (GHS), including depression{Melloh, 2013 #3620} while low capacity from pathology, injury, GHS or sedentary lifestyle increases the risk of LBP{Elfering, 2008 #1706}. The need to consider modifiable factors is supported by recent research{Marich, 2017 #3518} that confirmed the relationship between dynamic physical tests, self-reported LBP and reduced function{Grönblad, 1994 #3497;Grönblad, 1997 #3499}.

Methods – Test Activities
… Three exercises previously investigated, 'repeated sit-ups', 'repeated squats', and 'extension in lying' (EIL){Grönblad, 1994 #3497}, showed a positive correlation with LBP and were, consequently, included. The sustained squat and leg extension exercises, respectively require functional movement{Panjabi, 2003 #2457;Hoffman, 2013 #2560} and a predominantly isometric abdominal co-activation{Johnson, 2017 #3603}, which occur or simulate daily, occupational and sports activities{Zannotti, 2002 #3615}. Other exercises were considered but excluded, such as active spine flexion which has shown poor correlation with LBP{Sullivan, 2000 #830}.

Discussion
Previous research demonstrated a relationship between dynamic physical tests, self-reported LBP and reduced function{Grönblad, 1994 #3497}. However, such research has been neglected in recent decades{Melloh, 2015 #2928;Gabel, 2015 #2229;Sterud, 2013 #3521;Robinson, 2017 #3501} as focus shifted towards physiological and radiological findings{Dubois., 2016 #3520;Kohns, 2017 #3527} and biopsychosocial attributes{Machado, 2016 #3479}. Grönblad et al.{, 1994 #3497} showed three physical exercises (repetitive sit-ups, squats, and EIL) had a positive correlation with LBP. Our current study builds on this research as it expands the number of test exercises. It also shows a higher statistical correlation between physical exercise tests and LBP than found previously. These findings with robust effect sizes, and the 95% confidence intervals{Osborne, 2017 #3529}, demonstrate a substantial relationship. Our results indicate that for each increase in the exercise number accomplished, the odds of having some LBP were about one-third less than that of those participants accomplishing one fewer exercise.

• "Analysis of the findings may indicate what movements, or lack of movements, might be associated with the presence and/or absence of LBP. The outcomes may contribute to both the understanding of the relevance of functional movement and exercises as well as provide direction for future prospective studies..." My understanding is that numerous structures in the low back and nearby regions may be the sources of LBP. Thus, there is no 'one size fits all' intervention to prevent and treat LBP. Therefore, implementing the same preventive regime for all to prevent and treat LBP would be irrational.
The authors completely agree – there is no panacea. This is why we state that analysis may show a correlation between functional movements and LBP, their relevance and consequently a direction for future studies. Normal movement stresses normal tissue normally, and tight or weak structures are susceptible to overload from normal movement – this is the concept of the expanded Panjabi stability model {Hoffman, 2013 #2560}. The text is modified:
Once established, analysis of the findings might indicate what movements, or lack thereof, might be associated with the presence and/or absence of LBP for individuals in different occupational and physical activity settings. The outcomes might contribute to understanding the relevance of functional movement and exercises in relation to LBP, and provide a direction for future prospective studies. Such studies could identify specific functional movements for specific tasks or risk groups, then

provide structured exercise regimens that might reduce LBP and its predisposition.

3. METHODS

• There is no information about the inclusion and exclusion criteria of participants into the study.

The text is now modified to reflect this:

The test exercises were selected based on having significant elements of lumbo-pelvic-hip function and being recognized for reducing symptomology or risk of LBP. The five selected exercises were chosen to represent a balanced variation of functions required for normal daily activities{Hoffman, 2016 #3257}. Three exercises previously investigated, 'repeated sit-ups', 'repeated squats', and 'extension in lying' (EIL){Grönblad, 1994 #3497}, showed a positive correlation with LBP and were consequently included. The sustained squat and leg extension exercises, respectively require functional movement{Panjabi, 2003 #2457;Hoffman, 2013 #2560} and a predominantly isometric abdominal co-activation{Johnson, 2017 #3603}, which occur or simulate daily, occupational and sports activities{Zannotti, 2002 #3615}. Other exercises were considered but excluded, such as active spine flexion which has shown poor correlation with LBP{Sullivan, 2000 #830}.

All participants were volunteers and performed five functional movement exercises during an educational session with other attendees, supervised by the session leader, a Sports Physiotherapist Certified in McKenzie Manual Diagnostic Therapy. The instructions for exercise completion are detailed below. Intra-observer reliability for screening tests movement instruction is moderate-high{Carlsson, 2013 #3590}.

Class participant numbers ranged from 8-26, with a total sample of n=539. Only participants who consented were included. Data was excluded if there was insufficient demographic information. Consequently, the sample was reduced to a total of n=422, age 38.6±15.3 years, range 18-64 years, 67.3% male (see Table 1). Males were predominant in manual occupational roles including maintenance and construction, while females were predominant in carer and resource management including child care, community services, library and records roles.

.

• The rationale for including these exercises and the objective(s) of each exercise should be provided. Details about the criteria to identify those can and cannot perform each of five exercises should also be included.

The text is now modified to reflect this as detailed above. Also the section on Test Activities has been included as noted below / in Figure 1:

Test Activities

1) EIL: extension in lying, held for 3 seconds:

• Justification for inclusion: maximal lumbar extension simulates the physical properties of normal spinal movements{Panjabi, 1994 #3605;Hoffman, 2013 #2560} because limited extension{Steele, 2013 #3633} is related to LBP{Mazzone, 2016 #3631}, clinically impaired spinal control{Apeldoorn, 2016 #3635}, and may inhibit symptom centralization{McKenzie, 2003 #3599;Scannell, 2009 #3632}.

• Instructions to participants: lie face down, hands beneath shoulders, forehead on the floor. Keep your pelvis on the floor, breathe in, press with your arms, raise your chest off the ground, breathing out and increasing the movement till your arms are straight. Hold three seconds.

• Successful completion: hips/pelvis remain in contact with floor, arms fully extended.

• Test reliability: ICC=0.95-0.98{Youdas, 1995 #3601}.

2) SITUP: sit-up from supine with knees flexed and the arms passing the knees to or beyond the elbow whilst exhaling, performed 10 times:

• Justification: through range, active concentric and eccentric trunk flexion control enables the lumbar spine to dissipate and distribute load and provides a stable area for performing limb and trunk activities{Lehman, 2005 #3608;Abboud, 2017 #3609;O'Sullivan, 2005 #3256;Hoffman, 2013 #2560}.

• Instructions: lying face-up on the floor, knees bent, feet flat, arms straight and hands on thighs.

Breathe in, slowly sit-up whilst breathing out, move the elbows to touch your knees, rolling forward and up from the floor in a continuous movement, until everything above the buttocks is not touching the ground and your elbows reach your knees. Lower down in a continuous movement reversing the motion without falling or dropping while breathing out. Repeat 10-times.
• Completion: no sudden/rapid inertial motion, trunk not held rigid, feet remain on floor, elbows reach/pass the knees, body doesn't drop down.
• Reliability: ICC=0.995{Fry, 2015 #3600}.

3) LEGEXT: supine bilateral leg extension starting with the knees over the umbilicus, the legs then extending until the heels touch the ground with the knees at or near to full extension, performed 10 times:
• Justification: abdominal muscles are used predominantly isometrically to stabilize the body during this exercise{Johnson, 2017 #3603;Arokoski, 2004 #3636} and relevant to performing many household, occupational and sports activities{Zannotti, 2002 #3615}. The exercise provides co-activation significantly greater than in sit-ups/curl{Shields, 1997 #3616} enabling testing of rectus abdominis muscle and the internal and external oblique muscle activation{Johnson, 2017 #3603} reducing LBP risk when part of a motor control exercise program{Byström, 2013 #3637}.
• Instructions: lying on back on floor breathing in, head in contact or elevated, knees bent and above the umbilicus, lower back contacts the floor hands by side or under buttocks. Both legs are straightened, knees straightening until heels touch floor while breathing out. Small amounts of knee flexion are permitted. Return legs to the start position. Repeat 10 times.
• Completion: back and buttocks contact the floor, heels touch the ground, hands remain in start position.
• Reliability: [double] leg lowering (ICC=0.81-1.00){Zannotti, 2002 #3615} ICC=0.98{Enoch, 2011 #3614}; in active single leg raise ICC3.3=0.95-0.97{Linek, 2015 #3604}; abdominal muscle percentages "time active" is 54-86%{Johnson, 2017 #3603}.
4) SQUAT: 'toilet squat' - barefooted, feet and heels flat, hands touching feet, held for 3 seconds:
• Justification: squatting is frequently used and associated with many ADLs. It requires optimal lumbar flexion control to ensure normal spinal movements are maintained{Panjabi, 1994 #3605;Hoffman, 2013 #2560}, and shear-forces/lateral-movement are minimalized{Schoenfeld, 2010 #3610}. Squatting is a complex multi-segmental functional movement requiring coordinated biomechanical and neuromuscular components involving the leg and pelvic joints and muscles, respiratory system, with prime-mover muscle activation not significantly affected by common variations in kinetic chain continuity{Clark, 2012 #3612}. A semi-rigid spine eliminates planar motion but retains antero-posterior spinal integrity, as spinal flexion generally increases with hip flexion and the associated synergistic lumbar-pelvic action{Schoenfeld, 2010 #3610;Hsiang, 1997 #3611} which reduces the risk of LBP{Welch, 2015 #3634}.
• Instructions: stand comfortably, feet shoulder-width apart, arms loosely at your side. Breathe in, slowly squat, as though using a squat-toilet, allow the arms to move forward and hands touch the feet. Hold for 3 seconds.
• Completion: pelvis is lowered, heels/feet flat, fingers touch the feet.
• Reliability: Intra-rater Kappa=0.81-1.00 when tested alone{Edwards, 2017 #3595}; ICC>0.60 within a multi-exercise screen{Moran, 2016 #3593} and ICC=0.81{Bonazza, 2017 #3594}.

5) RISEUP: full squat and stand-up, performed 5 times with the head rising at the slightly before or at the same time as the buttocks.
• Justification: repeated squatting is functional and readily transfers to multiple ADLs. It requires coordinated prime-mover muscle activation and endurance{Clark, 2012 #3612} being the technique of choice for manual handling as net moments, muscle forces and internal spinal loads related to compression and shear force are reduced{Bazrgari, 2007 #3613}. Reduces LBP risk and is critical for normal spinal movement{Panjabi, 1994 #3605;Hoffman, 2013 #2560}.
• Instructions: complete the squat position described then rise to full standing with the head rising at

the slightly before or at the same time as the buttocks. Repeat five-times, a short rest is permitted.
• Completion: full squat action is completed as described; on rising the buttocks/pelvis do not rise before the trunk, i.e., no knee extension before hip extension.
• Reliability: ICC=0.61-0.80, standard error of measurement<3%{Rahmani, 2001 #3602}.

• Each participant also completed a questionnaire: 'How often do you have low back pain?'. What was the timeframe for the question, e.g. 1 year or lifetime?
This was deliberately left open ended as the variation in age and life experience of each participating individual was considerable. Consequently, it was described as within their own context and daily life.
The text is now modified:
During the educational sessions each participant completed a self-report questionnaire: 'How often do you have low back pain?' with three-response options: 'rarely/none', 'sometimes' or 'always/mostly', with the time frame and symptoms interpreted within their life context.

• "This 'LBP' group was, subsequently, further dichotomized into 'Some' and 'Most' to sub-categorize the severity of LBP being present in their lives 'sometimes' or 'most of/all the time' (Some/Most)…" I am a bit confused with this statement. What were the response options correlated with the number of exercises the individual was able to perform successfully?
The text is now modified to clarify this as follows:
This 3-point scale is condensed from the World Health Organisation's five-points: 'never', 'rarely', 'sometimes', 'often', and 'very often'{Kessler, 2005 #3542}. The three-point response provides an 'intermediate' option which is critical from psychological and statistical perspectives. Psychologically, three cognitive perspectives facilitate response accuracy by reducing cognitive load{Albarracin, 2005 #1729;Gabel, 2010 #1529} which improves precision and consistency{Krosnick, 1991 #1728}. Statistically, responses were coded on a 0-1-2 scale{Jacoby, 1971 #1621;Newcombe, 2001 #2598}: 0=rarely/none (No LBP), 1=sometimes (Some LBP), 2=always/mostly (Most LBP)..

• There is no information about the test-retest reliability of performing the exercises. Poor test-retest reliability would compromise the internal validity of the study.
The text is now modified to reflect this: - please see the edits above for Test Activities

• Details about statistical analysis are essentially required. Text is now modified as follows to reflect this:
Statistical analysis was performed using SPSS 23.0 for Windows with significance set at $p<0.05$. Following preliminary data screening to ensure data quality (e.g., no aberrant values), an initial cross-tabulation of LBP (lower back pain: 0=none, 1=some, 2=most) and number of exercises was performed to explore whether self-reported LBP was related to the number of exercises completed. A chi-square test evaluated whether the null hypothesis (that the number of exercises completed would be consistent across LBP groups) was tenable or able to be rejected.
A multinomial logistic regression was performed, exploring whether the number of exercises (EX_SUM) predicted LBP (categorized as 0, 1, 2) to test the null hypothesis that the probability or odds of being classified into LBP groups are not different because of number of exercises performed; and if rejected, to quantify the change in odds or probability of LBP as it relates to number of exercises performed. This test also allowed us to evaluate whether participant gender interacted with EX_SUM, or whether there were non-linear effects present. Regression diagnostics for this analysis (e.g., residuals, influence) were examined to ensure no aberrant cases were inappropriately influencing the analysis{Osborne, 2017 #3529}. None were identified.
Finally, if the null hypothesis from the prior multinomial logistic regression was rejected, we performed a second multinomial logistic regression on LBP entering each exercise as a predictor (rather than simply the count of number of exercises completed) to examine whether all exercises were uniquely

predictive or whether some subset of exercises were more predictive than others. All five exercises were entered simultaneously, allowing for examination of unique effects of each variable controlling for all other variables in the equation. Regression diagnostics were examined and no aberrant cases were identified{Osborne, 2017 #3529}.

4. RESULTS

• Demographic information of participants is essentially required.

The provision of the demographic data of age, gender, and occupation is presented in Table 1.

• How about the relationship between each exercise and the presence and/or absence of LBP? This information, similar to Table 1, would be helpful to understand the results in Table 4 and 5.

The text is now modified to reflect this – Table 4 -5 are now Table 5 -6:

A second multinomial logistic regression with the five exercise variables entered individually, rather than entering the total number accomplished, evaluated whether tests were individually predictive of LBP. As shown in Table 5, overall the effect was similarly strong{Osborne, 2017 #3529} (initial-2LL=429.93, final-2LL=147.40, $X2(2)=282.53$, $p<0.001$). As Table 6 presents, most exercises were individually predictive of LBP (when LBP=1, EIL was not uniquely predictive with all other variables in the equation). All others were statistically significant ($p<0.002$) with odds ratios ranging in magnitude from 0.21 to 0.38. For "Most" LBP (LBP=2), all exercises were significant independent predictors of LBP (all $p<0.017$), with odds ratios ranging from 0.09-0.35.

Sensitivity for the first analysis (percent of participants with LBP correctly classified into LBP category) was 82.33%, and specificity (percent of participants with no LBP classified as such) was 85.55%. The positive predictive value (true positives divided by true and false positives) was 89.13%; and negative predictive value (true negatives divided by true and false negatives) was 77.08%. Sensitivity for the second analysis was 79.52%, and specificity was 87.86%. Positive predictive value was 90.41%, and negative predictive value was 74.88%.

• "…less than 3% of all participants were able to complete one or no exercises…" This sentence is quite difficult to comprehend.

The text is now modified to reflect this:

For descriptive purposes, a cross-tabulation of LBP (0=none, 1=Some, 2=Most) and the number of exercises accomplished is presented in Table 2. Most participants reporting no LBP could complete most exercises. For individuals with no LBP, 85.5% could complete at least four exercises. Exercise completion dropped significantly for participants with "Some" LBP. In this group, only 22.9% were able to complete four or more exercises, and for participants with "Most" LBP, only 10.5% were able to complete four or more exercises. Analyzing participants in each category who failed to complete more than one exercise, the pattern is reversed. Only 2.9% of those with no LBP had trouble completing more than one exercise, while 23.7% of those with "some LBP" and 74.3% of those with "most LBP" were unable to complete more than one. A Pearson Chi-square test was performed demonstrating a significant relationship between the variables of 'LBP' and 'number of exercises performed' ($X2(10)=300.61$, $p<0 .001$).

• It is difficult to understand the findings because no information about statistical analysis is provided. For example, no information about confounders included in the logistic regression analysis and how they were derived is available. Why was the gender effect tested only?

The text is now modified to reflect this:

In Methods as noted and detailed above: -

Next, a multinomial logistic regression was performed, exploring whether the number of exercises (EX_SUM) predicted LBP (categorized as 0, 1, 2). While this …

and

Finally, if the null hypothesis from the prior multinomial logistic regression was rejected, we performed

a second multinomial logistic regression on LBP entering each …

and further in the results section:
A multinomial logistic regression predicting LBP (0, 1, 2, with 0 being the reference group) from the count of exercises that could be completed (EX_SUM, ranging from 0-5), showed …
and
A second multinomial logistic regression entered the five exercise variables individually, rather than entering the total number accomplished, to test whether particular tests were individually diagnostic (predictive) of LBP. As shown in Table 5, overall ….

Also we have added a final paragraph as follows:
We also took in to consideration a simple analysis relating the presence or absence of LBP to exercises. This approach, combining two groups of LBP (some, mostly) into one category potentially reduces the goodness of the analysis by combining two different groups into one heterogeneous group. If the two groups were distinct, this would increase error variance and decrease the power and informativeness of the analyses. Ancillary binary logistic regression analyses therefore tested the null hypothesis that the two LBP groups were similar. Results of this analysis, which predicted LBP (i.e., some vs. mostly) showed that EX_SUM was significantly related to this outcome (initial-2LL=339.05, final-2LL=284.96, X2(1)=54.09, p<0.001), leading us to reject the null hypothesis and assert that these two groups are significantly distinct, and therefore inappropriate for combining{Osborne, 2017 #3529}.
We have also made a note in the study limitations as follows:
Another potential limitation was that participant self-reported gender was the only potential moderator or confounding variable included in the data. As noted above, gender itself was not a significant predictor in any analysis (p>0.80), and thus not included in analyses reported. We were unable to test for a significant interaction between gender and exercises (e.g., EX_SUM) due to quasi-complete separation in the data. However, a trend appeared where the effects for males were slightly stronger. This might represent a direction for future research within larger samples, or simply a sample artefact.
And in Future research
Furthermore, this study had limited demographic variables. Consequently, future research may consider moderating factors aside from gender. Perhaps age is a differential consideration. However, the very strong analyses effects observed and that our lack of explicitly modeling these hidden variables would have biased the results toward the null, it is unlikely that unobserved variables are true confounders, but might clarify and increase the sensitivity of some effects if modeled. As an observational study, however, it was not possible to indicate whether gradually training individuals to complete these five exercises could facilitate reductions in LBP. From the several authors clinical management protocol it can be speculated that this appears possible.

5. DISCUSSION
• “These exercises could be used clinically to diagnose the potential severity of LBP, and perhaps severity of impairment” This sentence was not supported by the findings of this study, i.e. no data collection regarding pain severity or impairment.
The text is now modified to reflect this:
Consequently, these exercises have the potential to be investigated in future research in terms of the ability to provide a clinical diagnosis related to the potential or risk that an individual may development LBP, and perhaps even future impairment.

• “Physical functional tests, especially those emphasized in this study, are directed primarily toward the abdominal and lumbo-pelvic muscles and their coordinated activity” LBP in different occupations is unlikely to originate from identical causes because patients are exposed to different risk factors. In

this case, is it possible that lumbar stability may not be the problem for everyone? Some discussion to acknowledge this fact would be useful.
We appreciate this input from the reviewer and the text is now modified to reflect this with addition to the paragraph and modification of the existing one:
Additional paragraph:
Exercise therapy is an efficient, cost-effective LBP management strategy{van Middelkoop, 2010 #3638;Lin, 2011 #3639} but there is no evidence to support any single exercise. Coordinated muscle activity around the lumbo-pelvic region is considered vital for mechanical spinal stability{McGill, 2003 #3640;Hoffman, 2013 #2560}. Several rehabilitative "stabilization exercise" approaches emphasize retraining functional movement patterns, rather than focusing on specific muscles{Ikeda, 2012 #3642;Bell, 2009 #1668;Hoffman, 2016 #3257}. The tests we chose activate and challenge the global muscles of the abdomen and trunk, the "abdominal brace" mechanism{Myrtos, 2012 #3643}, and their ability to act and interact in a synergistic and functional manner. We screened functional test performance where the aim was assessing participants' functional status regardless or not of LBP and its known or potential cause. AS LBP continues to increase in industrial societies with no clear cause it is important to consider risk factors of physical workload and awkward posture{Machado, 2016 #3479} as well as preventative strategies that may play a key role in reducing health care system demands and societal support. The exercise tests we used primarily address abdominal and lumbo-pelvic muscles, and their coordination with lower limb muscle activity and maintenance of balance. This coordination was recently defined as 'integral' in understanding lumbar stability as a complex integrated model{Hoffman, 2013 #2560}. Personal efficiency in physical self-test completion can act as a screening methodology for individuals at risk of LBP. It is, however, important that the method of test performance is considered e.g., there is no relation demonstrated between sit-up performance and LBP when the feet are held{Jackson, 1998 #3644}. This action preferences hip flexor activity over abdominal participation. Alternative actions that preference abdominal muscles, e.g. partial curl-up, are more highly correlated to LBP{Parfrey, 2008 #3551;Moya-Ramón, 2017 #3552}. Our results provide guidance for future work that may contribute to a comprehensive screening, prevention and management approach to LBP.

• I am not sure that the authors provided possible explanations or hypotheses for why the number of exercises completed might predict the presence and/or absence of LBP in the sample of population.
The text is now modified to reflect this:
This observational study investigated council workers, as an implied representative general working population sample, and evaluated whether the ability, or not, to perform five back-related exercises could determine or predict the presence or absence of LBP. We hypothesized that the test-exercises would demonstrate the ability of the lumbar spine to: move in a controlled manner through normal range as a complex multi-segmental functional activity with coordinated biomechanical and neuromuscular components; and be stabilized, as part of the lumbo-pelvic-hip complex, through motor control of the integrated muscular system{Panjabi, 2003 #2457;Hoffman, 2013 #2560}. Consequently, the ability to perform the exercises would correlate with lower self-reported LBP. Once established, analysis of the findings may …

• "Strengths of this study include the prospective nature…" Is this study cross-sectional in design?
Correct, this error has been rectified. The text is now modified to reflect this:
The strengths of this study include the cross-sectional nature, the sample including both genders, diverse age groups and occupations but within one organisation and geographical region.

• "…with diverse age groups and occupations…" and "…degree of homogeneity in the sample…" seems to contradict each other.
The text is modified to reflect this
… This enabled continuity, and a degree of homogeneity in the otherwise varied sample, that

strengthened the statistical findings with respect to general working populations.

• One of the limitations of the study is its cross sectional design which does not allow for the causal relationship to be investigated. Also, a subjective diagnosis of LBP is another study weakness.
The text is modified to reflect this: see edits above for Test Activities and below for limitations/weakness:
… Causality however, cannot be inferred from this study.
Other exercise tests may have similar utility. In choosing the exercise tests, we did not consider exercise dose and specificity for age and gender and these may be confounding factors. However, the statistical findings showed that the exercises chosen were relevant and that neither gender nor age influenced the results.
Another potential limitation was that participant self-reported gender was the only potential moderator or confounding variable included in the data. As noted above, gender itself was not a significant predictor in any analysis ($p>0.80$), and thus not included in analyses reported. We were unable to test for a significant interaction between gender and exercises (e.g., EX_SUM) due to quasi-complete separation in the data. However, a trend appeared where the effects for males were slightly stronger. This might represent a direction for future research within larger samples, or simply a sample artefact

6. Others
•Below Table 1 should add the meaning of 1, 2, and 3.
See modified Table 2 as follows with the heading below each number:
0
None 1
Some 2
Most

• Would it be easier to comprehend Table 3 and 5 by modifying the current (statistical) format to simpler format?
The original Tables 3 and 5, [now 4 and 6] are unchanged under the bio-statisticians advice as they are presented in the standard statistical format. This ensures the greatest precision on the data analyzed. The concern to simplify further would necessitate leaving out some important information such as the confidence intervals.

Responses to Reviewer: 2
Reviewer Name: Achim Elfering
Institution and Country: University of Bern, Switzerland Please state any competing interests: None declared

Please leave your comments for the authors below

Bmjopen-2017-020946
Theory

The ms reports an inverse association between the capacity to perform 5 exercises and self-reported frequency of back pain. Data are cross-sectional. The association is worth reporting –

However the author(s) should be aware that causality cannot be inferred. Causality would need a temporal sequence of antecedent and consequence and causality would need to rule out alternative explanations.
The text is now modified as follows to reflect this within the limitations section:
"… Causality however, cannot be inferred from this study.
Other exercise tests may have similar utility. In choosing the exercise tests, we did not consider exercise dose and specificity ….."

The author(s) acknowledge this shortcoming but sometimes there is also some « causal » writing , e.g., in admitting a „prospective nature" of the study on page 7 line 58.
The text is now modified as follows to reflect this as follows:
"The strengths of this study include the cross-sectional nature, …"

Indeed there is evidence that back pain reduces capacity and low capacity increases risk for back pain and third variables (e.g., fittness, health status, comorbidities) may affect both capacity and back pain.
The author(s) should consider such models in more depth in the introduction and in the discussion.
This aspect has been expanded upon within the introduction with modifications to the text as follows:
Low back pain (LBP) is among the world's most prevalent occupational disorders in working populations{Deyo, 2001 #3492} and major global public-health concerns{Balagué, 2012 #3618}, affecting 12 percent of the world's population at any given time{Murray, 2013 #3476;Balagué, 2012 #3618} with lifetime prevalence at 84% and chronicity around 23%{Balagué, 2012 #3618}. …

LBP disorders are multi-factorial with individual symptomology influenced by various patho-anatomical, physical, neuro-physiological, psychological and social contributors{O'Sullivan, 2005 #3256;Hoffman, 2013 #2560}. Consequently, voluntary activities that involve lumbo-pelvic specific exercises are effective in primary and secondary LBP prevention{Broonen, 2011 #3607}. Such exercises improve fitness and occupational status by diminishing disability and problem severity{Henchoz, 2008 #3606;Hoffman, 2016 #3257} and may counter selective atrophy of Type II fibers found in the presence of pathological changes{Ng, 1998 #3617;Goubert, 2016 #3621}. However, muscle recruitment remains predominantly neural-based during rehabilitation with psychological adaptations derived from improved motivation and pain tolerance{Mannion, 2001 #3626}. The conundrum remains that LBP reduces functional capacity, fitness and general health status (GHS), including depression{Melloh, 2013 #3620} while low capacity from pathology, injury, GHS or sedentary lifestyle increases the risk of LBP{Elfering, 2008 #1706}. The need to consider modifiable factors is supported by recent research{Marich, 2017 #3518} that confirmed the relationship between dynamic physical tests, self-reported LBP and reduced function{Grönblad, 1994 #3497;Grönblad, 1997 #3499}.

It would be interesting to control for potential third variables (e.g., fittness, etc. ) in analyses. It would also be interesting to see whether exercises are uniquely related to back pain over and above self-report screening information from questionnaires. If author(s) have the information they should add analyses to the ms.
We thank the reviewer and agree that this would be a beneficial addition to the manuscript, however we do not have this information and it cannot be gained retrospectively. Consequently, it cannot be included.

Methods
Please report the participation rate: How many participants of educational classes agreed to do the exercises?
The text is now modified to reflect this:
Class participant numbers ranged from 8-26, with a total sample of n=539. Only participants who consented were included. Data was excluded if there was insufficient demographic information. Consequently, the sample was reduced to a total of n=422, age 38.6±15.3 years, range 18-64 years, 67.3% male (see Table 1). Males were predominant in manual occupational roles including maintenance and construction, while females were predominant in carer and resource management including child care, community services, library and records roles.

Please report the context of exercise.
What was the instruction ?
The text is now modified to reflect this: - please see the edits above for Test Activities – Figure1

Which criteria were chosen to define an exercise to be done or failed ?
The text is now modified to reflect this: - please see the edits above for Test Activities – Figure1
Please report reliability estimates on these criteria.
The text is now modified to reflect this: - please see the edits above for Test Activities – Figure1

Were the exercise done alone or in presence of other class members ?
All exercises were completed in the presence of all class attendees. The text is modified as follows:
All participants were volunteers and performed five functional movement exercises during an educational session with other attendees, supervised by the session leader, a Sports Physiotherapist Certified in McKenzie Manual Diagnostic Therapy. The instructions for exercise completion are detailed below. Intra-observer reliability for screening tests movement instruction is moderate-high{Carlsson, 2013 #3590}.

Were the exercise done before or after filling out the questionnaire on low back pain ?
The exercises were performed in mid-session and after questionnaire completion. The text is modified:
During the educational sessions each participant completed a self-report questionnaire: ‘How often do you have low back pain?’ with three-response options: ‘rarely/none’, ‘sometimes’ or ‘always/mostly’, with the time frame and symptoms interpreted within their life context. This 3-point scale is condensed from the World Health Organisation’s five-points …

Why did the author(s) cite a questionnaire on attention-deficit/hyperactivity disorder (ref 55) for their own question on back pain?
This reference by {Kessler, 2005 #3542}, and the questionnaire it referred to, was cited for its five-point response option, not for its content or subject matter. The five-point response option is a response option template that complies with the WHO standards. It can be reduced or modified to provide a three-point scale by combining or pooling the lower and upper options as noted below: 0=rarely/none (0-3), 1=sometimes (4-6), 2=always/mostly (7-10) . This three-point or tri-chotomous response option, as used in this study, has significant support in terms of the accuracy of the individual responses as a consequence of the reduced cognitive load to responding individuals {Krosnick, 1991 #1728;Albarracin, 2005 #1729}.

From Kessler 2005, WHO ASDS Self report scale:
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.424.7557&rep=rep1&type=pdf:
The response options are: never, rarely, sometimes, often, and very often. Patients were asked to answer the questions using a 6-month recall period.
0-1= never, 2-3=rarely, 4-6=sometimes, Often= 7-8, very often.= 9-10

Consequently our Study used the following:
0 = None = rarely/none (0-3)
ie no LBP ever or an occasional (easily forgotten and <4 times per year / every 3 months or quarter or normal season), brief (lasting <1 day), minimally symptomatic or functionally inhibiting bout that caused minimal or no disability;
1 = Some = sometimes (4-6)
o ie LBP that was not easily forgotten (or >4 times per year / every 3 months or quarter or normal season; eg monthly or more), not brief (lasting >1 day), symptomatic and functionally inhibiting bout that caused some disability or interruption of daily life
2 = Most = always/mostly (7-10)
o ie Not easily forgotten, occurring frequently as in daily to weekly, not brief (lasting >1 day), symptomatic & functionally inhibiting bout causing some disability or interruption of daily life.
The text has been modified to reflect this, and additional references provided to support both aspects of the preferred three point option and that of cognitive demand or load. The text is now modified as follows:
This 3-point scale is condensed from the World Health Organisation's five-points: 'never', 'rarely', 'sometimes', 'often', and 'very often'{Kessler, 2005 #3542}. The three-point response provides an 'intermediate' option which is critical from psychological and statistical perspectives. Psychologically, three cognitive perspectives facilitate response accuracy by reducing cognitive load{Albarracin, 2005 #1729;Gabel, 2010 #1529} which improves precision and consistency{Krosnick, 1991 #1728}. Statistically, responses were coded on a 0-1-2 scale{Jacoby, 1971 #1621;Newcombe, 2001 #2598}: 0=rarely/none (No LBP), 1=sometimes (Some LBP), 2=always/mostly (Most LBP).

Please report the sensitivity, specificity and predictive value of exercises.
The text is now modified to reflect this:
A multinomial logistic regression predicting LBP (0, 1, 2, with 0 being ….

A second multinomial logistic regression with the five exercise variables entered individually, rather than entering the total number accomplished, …

Sensitivity for the first analysis (percent of participants with LBP correctly classified into LBP category) was 82.33%, and specificity (percent of participants with no LBP classified as such) was 85.55%. The positive predictive value (true positives divided by true and false positives) was 89.13%; and negative predictive value (true negatives divided by true and false negatives) was 77.08%. Sensitivity for the second analysis was 79.52%, and specificity was 87.86%. Positive predictive value was 90.41%, and negative predictive value was 74.88%..
Also we noted as follows:
We also took in to consideration a simple analysis relating the presence or absence of LBP to exercises. This approach, combining two groups of LBP (some, mostly) into one category potentially reduces the goodness of the analysis by combining two different groups into one heterogeneous group. If the two groups were distinct, this would increase error variance and decrease the power and informativeness of the analyses. Ancillary binary logistic regression analyses therefore tested the null hypothesis that the two LBP groups were similar. Results of this analysis, which predicted LBP (i.e., some vs. mostly) showed that EX_SUM was significantly related to this outcome (initial-2LL=339.05, final-2LL=284.96, $X2(1)=54.09$, $p<0.001$), leading us to reject the null hypothesis and assert that these two groups are significantly distinct, and therefore inappropriate for combining{Osborne, 2017 #3529}.

## VERSION 2 – REVIEW

| | |
|---|---|
| **REVIEWER** | Prawit Janwantanakul |

| | |
|---|---|
| | Department of Physical Therapy, Faculty of Allied Health Sciences, Chulalongkorn University, Bangkok, Thailand |
| **REVIEW RETURNED** | 11-Apr-2018 |

| | |
|---|---|
| **GENERAL COMMENTS** | None. |

| | |
|---|---|
| **REVIEWER** | Achim Elfering<br>University of Bern, Switzerland |
| **REVIEW RETURNED** | 16-Apr-2018 |

| | |
|---|---|
| **GENERAL COMMENTS** | The author(s) sufficiently addressed my points. |