

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Differences in scapular upward rotation, pectoralis minor and levator scapulae muscle length between the symptomatic, the contralateral asymptomatic shoulder and control subjects: A cross-sectional study in a Spanish primary care setting.
AUTHORS	Navarro-Ledesma, Santiago; Fernandez-Sanchez, Manuel; STRUYF, FILIP; Martinez-Calderon, Javier; Miguel Morales-Asencio, Jose; LUQUE-SUAREZ, ALEJANDRO

VERSION 1 - REVIEW

REVIEWER	Mark Timmons Marshall University Huntington WV USA
REVIEW RETURNED	13-May-2018

GENERAL COMMENTS	<p>Thank you for asking me to review this paper. The paper addresses an important topic and has promise. I, however, have concerns regarding the methods and the statistical analysis. I have made several comments on the attached pdf. Some of my concern might be addressed by simply rewriting the paper for clarity and reorganizing the paper. If the authors can address my concerns I am happy to review the paper again.</p> <p>The reviewer provided a marked copy with additional comments. Please contact the publisher for full details.</p>
-------------------------	---

REVIEWER	Dayana Patricia Rosa Universidad Industrial de Santander - Bucaramanga, Colombia
REVIEW RETURNED	01-Jun-2018

GENERAL COMMENTS	<p>Paper Revision</p> <p>Abstract Please add the values of MDC and mean difference in the Abstract, making it more easy to understand by clinicians. The values presented in the conclusion was not described in the results section.</p> <p>Introduction Recent evidence challenges the concepts behind subacromial impingement (references below). These studies have suggested to abandon the "SIS" diagnostic term, used by the present study. Please be more clear about which population did you include: individuals with subacromial impingement syndrome or symptoms of impingement.</p>
-------------------------	---

1. Braman JP, et al. Shoulder impingement revisited: evolution of diagnostic understanding in orthopedic surgery and physical therapy. *Med Biol Eng Comput.* 2014;
2. Lawrence RL, Effect of glenohumeral elevation on subacromial supraspinatus compression risk during simulated reaching. *J Orthop Res.* 2017;
3. Bey MJ, et al. In vivo measurement of subacromial space width during shoulder elevation: technique and preliminary results in patients following unilateral rotator cuff repair. *Clin Biomech.* 2007.
4. Ludewig PM, Lawrence RL, Braman JP. What's in a name? Using movement system diagnoses versus pathoanatomic diagnoses. *J Orthop Sports Phys Ther.* May 2013;43(5):280-283.

The Lewis and Valentine (2007) study's should be considered in the Introduction and Discussion Section, considering that they used the acromion table distance to estimate the PM length. The authors should revise the papers indicated below to strength the introduction and discussion section:

1. Rosa DP, Borstad JD, Pires ED, Camargo PR. Reliability of measuring pectoralis minor muscle resting length in subjects with and without signs of shoulder impingement. *Braz J Phys Ther.* Mar 15 2016;20(2):176-183.
2. Rosa DP, Borstad JD, Pogetti LS, Camargo PR. Effects of a stretching protocol for the pectoralis minor on muscle length, function, and scapular kinematics in individuals with and without shoulder pain. *J Hand Ther.* Jan - Mar 2017;30(1):20-29.
3. Rosa DP, Rodrigo V. Santos, Gava V, Borstad JD, Camargo PR. Shoulder external rotation range of motion and pectoralis minor length in individuals with and without shoulder pain. *Phys Theory Pract.*

The SPADI questionnaire was not described as a aim of the study, but was described as a result.

Methods

Please add references to inclusion criteria.

Are you sure that an SPADI score of 15 point is considered asymptomatic shoulder?

Delete the reference Ekeberg et al., 2010 inside the manuscript, please.

Please add reference for the PM length measurement adopted.

Why did you evaluate the muscle length in supine and not in the upright position as described by Borstad and Ludewig (2005)?

Considering the statistical analysis adopted, should not the authors adjust the alfa based on the number of comparisons?

I think that the ICC, SEM, and MDC results should be in a table.

Discussion

Considered the affirmation that "this was the first study to compared SUR between symptomatic and asymptomatic subjects" Actually, it was the first study to compared SUR, using the method described. Many previous studies used 3d analysis.

The authors should revise the manuscript: Haik MN, Alburquerque-Sendín F, Camargo PR. Reliability and minimal detectable change of 3-dimensional scapular orientation in individuals with and without shoulder impingement. *J Orthop Sports Phys Ther.* May 2014;44(5):341-349, which report ICC, SEM and MDC values for the upward rotation with an 3D-analysis. The ultrasound clinical assessment was not described in the methods section.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Mark Timmons

Institution and Country: Marshall University Huntington WV USA

Please state any competing interests or state 'None declared': None

Please leave your comments for the authors below

Thank you for asking me to review this paper. The paper addresses an important topic and has promise. I, however, have concerns regarding the methods and the statistical analysis. I have made several comments on the attached pdf. Some of my concern might be addressed by simply rewriting the paper for clarity and reorganizing the paper. If the authors can address my concerns I am happy to review the paper again.

Response: Thank you very much for your time and effort in reviewing this paper. And thanks for your positive feedback. Please, find below a document in response point by point to your comments.

INTRODUCTION

Response 1: This typo has been corrected.

Response 2: This has been changed, and now it appears as follows. "...there is a need to explore different non-invasive strategies in these patients."

Response 3: This has been modified, and now it appears as follows: "To date, there is inconsistent evidence to support a relationship between SAPS symptoms and scapular orientation".

Response 4: Regarding the definition of PMI and LSI, we have decided to generally mention these muscles rather than specific measurements that assess their length at this point of the introduction, as follows: "The pectoralis minor and levator scapulae muscle length (LSI) (Borstad JD, Ludewig PM, 2005; Lee et al, 2016) have been traditionally assessed as their shortening may potentially influence scapular positioning". A definition and a more detailed description of pectoralis minor index (PMI) and levator scapulae index (LSI) can be found during the method section.

Regarding your suggestion of adding other muscles that can contribute to SAPS, as this manuscript is focused on muscles directly attached to scapula, we have incorporated this information to the introduction: "Specifically, both a decreased activation and strength of serratus anterior, as well as alterations in upper trapezius/lower trapezius couple force, can alter scapular upward rotation and posterior tilt (Kibler et al, 2013). Likewise, pectoralis minor and levator scapulae muscles (LSI) (Borstad JD, Ludewig PM, 2005; Lee et al, 2016), and biceps short head (Kibler et al, 2013) have been traditionally assessed as their shortening may potentially influence scapular positioning."

Response 5: Thank you. In one hand, we have changed this information in the text, regarding the relationship between a reduced both scapular upward rotation and scapular posterior tilt. Now it appears as follows: "With regard to patterns of movement, there is conflicting evidence. While some studies have shown association between a reduced both scapular upward rotation (SUR) and scapular posterior tilt in SAPS (Struyf et al, 2011; Ellenbecker et al, 2010), others did attain inconclusive findings (Ratcliffe et al, 2014; Timmons et al, 2012)."

Response 6: In the last paragraph of the introduction section, we have included the following statement: "The null hypothesis (H0) was that there are no differences in these three different tests between groups. The alternative hypothesis (Ha) was that there are significant differences in these three tests between groups."

METHOD

Response 1: Yes, we do agree. We have moved this information to the PMI description.

Response 2: Thank you. This was a typo and has been corrected as suggested.

Response 3: Yes.

Response 4: Thank you. This information has been included in the text.

Response 5: This has been specified, as follows: "...as previously described by Borstad et al."

Response 6 and 7: Thanks for your interesting comments at this point. In one hand we agree that information related to reliability for SUR, PMI and LSI needs to be reported during the method section. We did not do that because, as this was not a purpose of this study, we decided to mention it during the results section. However, and based on another suggestion in the results section, we have moved this information to the data analysis section (method section).

Response 8: Thank you very much. This is a very good point. In one hand, we totally agree with the fact that a global statistical analysis that assesses whether there are differences between the three groups, is necessary. However, ANOVA test would not be appropriate, as there is not full independence among the three groups of the study. Symptomatic and contralateral shoulders belong to the same subject and they are not independent. Based on that, we have run the suggested statistical analysis using the Kruskal-Wallis test. The attained results have been incorporated to the results section, and all the manuscript has been adapted to these results.

RESULTS

Response 1: This has been changed, as follows: "Demographic characteristics are shown in table 1".

Response 2 and 3: As described aforementioned, we have moved this information to the data analysis section. Furthermore, we have added information regarding the intra-rater reliability, as follows: "Although it was not a purpose of this study, we calculated the intra-rater reliability for all the outcome measurements by using the intraclass correlation coefficient (ICC), in order to determine the minimal detectable change at 95% (MDC95)".

Response 4: This was a typo. We have amended it in the new version of this manuscript.

Response 5 and 6: Thank you. We agree that in the current form the results section looks a bit confusing. For that, we have made the following changes: (i) we have incorporated ICC and MDC values to Table 2; (ii) we have rewritten some subheadings; (iii) we have made some changes due to the new statistical analysis. We think these changes have improved the reading and comprehension of this part of the manuscript.

Regarding MDC95 for SUR, yes, this value is related to "degrees". This has been clarified in Table 2.

DISCUSSION

Response 1: This has been changed, as follows: "On the other hand..."

Response 2: The discussion has been adapted, basically the first paragraph Based on the new analysis that we have carried out. The following information has been included: "We found statistical significant differences between the three groups in SUR at 45, 90 and 135 degrees of shoulder elevation. Specifically, a decreased SUR in symptomatic shoulder compared to contralateral asymptomatic shoulder at 45 degrees, was achieved. When comparing symptomatic and control participants, an increased SUR at all positions (45, 90 and 135 degrees) was found in the symptomatic shoulders. Regarding PMI and LSI, there were not significant differences between all the groups."

Reviewer: 2

Reviewer Name: Dayana Patricia Rosa

Institution and Country: Universidad Industrial de Santander - Bucaramanga, Colombia

Please state any competing interests or state 'None declared': None declared

Response: Thank you for your time and effort in reviewing this paper. Please, find below responses to your comments.

Please leave your comments for the authors below

Paper Revision

Abstract

Please add the values of MDC and mean difference in the Abstract, making it easier to understand by clinicians.

The values presented in the conclusion were not described in the results section.

Response: Thank you. This has been corrected based on your suggestions.

Introduction

Recent evidence challenges the concepts behind subacromial impingement (references below). These studies have suggested to abandon the "SIS" diagnostic term, used by the present study. Please be more clear about which population did you include: individuals with subacromial impingement syndrome or symptoms of impingement.

1. Braman JP, et al. Shoulder impingement revisited: evolution of diagnostic understanding in orthopedic surgery and physical therapy. *Med Biol Eng Comput.* 2014;
2. Lawrence RL, Effect of glenohumeral elevation on subacromial supraspinatus compression risk during simulated reaching. *J Orthop Res.* 2017;
3. Bey MJ, et al. In vivo measurement of subacromial space width during shoulder elevation: technique and preliminary results in patients following unilateral rotator cuff repair. *Clin Biomech.* 2007.
4. Ludewig PM, Lawrence RL, Braman JP. What's in a name? Using movement system diagnoses versus pathoanatomic diagnoses. *J Orthop Sports Phys Ther.* May 2013;43(5):280-283.

Response: Thank you. This is an interesting point. We do agree with your comment and the evidence provided, not only in the studies aforementioned, but also in:

- Lewis 2016 (Rotator cuff related shoulder pain: Assessment, management and uncertainties)
- Lewis et al, 2015 (Rotator Cuff Tendinopathy: Navigating the Diagnosis- Management Conundrum)
- Lewis, 2015 (Bloodletting for pneumonia, prolonged bed rest for low back pain, is subacromial decompression another clinical illusion?)
- Beard et al, 2017 (Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial)

- Paavola et al, 2018 (Subacromial decompression versus diagnostic arthroscopy for shoulder impingement: randomised, placebo surgery controlled clinical trial)
- Lewis, 2018. The End of an Era?

Additionally, there is a tendency to avoid terminology because it could negatively impact on patients expectations of recovery, e.g., "impingement".

Based on all of this, we have consistently made reference to subacromial pain syndrome (SAPS) across the whole manuscript. In addition to this, in the second paragraph of the introduction section, we have added the following information: "It is defined as a non-traumatic, usually unilateral, shoulder problem that causes pain localized around the acromion, often worsening during or subsequent to lifting of the arm. (Diercks et al, 2014)". Moreover, we have included in the inclusion criteria references of Diercks et al, 2014 (SAPS), and Cools et al (2008) to support this point.

The Lewis and Valentine (2007) study's should be considered in the Introduction and Discussion Section, considering that they used the acromion table distance to estimate the PM length.

Response: This study had been included during the introduction in the previous version of this manuscript, although without an adequate description. Because of this, the following statement has been included in the third paragraph of the introduction: "Previous studies have reported normative values on pectoralis minor length in the dominant and non-dominant side in both symptomatic and control populations, by using the pectoralis minor index (Struyf et al, 2014), and the acromion-table distance test (Lewis et al, 2007)". As Lewis and Valentine (2007) study did not report differences between shoulder pain group and asymptomatic group, and due to they did not use the same procedure to assess pectoralis minor length as we used in the present study, we did not make reference to it during the discussion.

The authors should revise the papers indicated below to strength the introduction and discussion section:

1. Rosa DP, Borstad JD, Pires ED, Camargo PR. Reliability of measuring pectoralis minor muscle resting length in subjects with and without signs of shoulder impingement. *Braz J Phys Ther.* Mar 15 2016;20(2):176-183.
2. Rosa DP, Borstad JD, Pogetti LS, Camargo PR. Effects of a stretching protocol for the pectoralis minor on muscle length, function, and scapular kinematics in individuals with and without shoulder pain. *J Hand Ther.* Jan - Mar 2017;30(1):20-29.
3. Rosa DP, Rodrigo V. Santos, Gava V, Borstad JD, Camargo PR. Shoulder external rotation range of motion and pectoralis minor length in individuals with and without shoulder pain. *Phys Theory Pract.*

Response: Thank you very much for the provided information. These outstanding papers have contributed to increase the body of knowledge in the field. Thus, we have added to the following information: “Recently, pectoralis minor length and its shortening have received remarkable empirical attention, in terms of reliability study (Rosa et al, 2016), association with shoulder external rotation(Rosa et al, 2018), and as an outcome measurement after a stretching program in participants with shoulder pain(2016). However, differences between symptomatic groups and healthy controls were not calculated.”

The SPADI questionnaire was not described as a aim of the study, but was described as a result.

Response: We described SPADI mean values for both patients and healthy groups in Table 1 as part of the sample characteristics. As we did not carry out any analysis with shoulder pain and function data we believe that this does not merit to be considered as a purpose of this study.

Methods

Please add references to inclusion criteria.

Response: In the original draft a reference was included in the third inclusion criterion (Cools et al, 2008). Furthermore, we have included the reference of Diercks et al (2014) that defines SAPS as mentioned above. We believe that these two references support the inclusion criteria used in this study.

Are you sure that an SPADI score of 15 point is considered asymptomatic shoulder?

Response: This is a good point. We strongly agree that there is not a consensus on considering a determined SPADI score to conclude that an individual is “free of shoulder pain”. However, it is also true that we can not consider “painfree” individuals those who respond “NO” to the question: Have you got any shoulder pain at this moment? Based on that, we used the minimal clinically detectable change (MCDC) (Engelbrechtsen et al, 2010) to select the healthy controls. Additionally, we have added the following statement as a limitation of this study: “Lastly, including healthy controls by using a SPADI score below 15 points could mean bias.”

Delete the reference Ekeberg et al., 2010 inside the manuscript, please.

Response: Thank you. This was a typo and, thus, has been removed from the manuscript.

Please add reference for the PM length measurement adopted. Why did you evaluate the muscle length in supine and not in the upright position as described by Borstad and Ludewig (2005)?

Response: Regarding the PMI calculation, we followed the study of Borstad and Ludewing, 2005, as cited in the previous version of this manuscript. About the position adopted by the patient we followed the study of Mackenzie et al, 2016, as cited in the previous version. This position has been used in previous studies (Struyf et al, 2014; Cools et al, 2010), in order to minimize postural influences of the thoracic spine and to optimize the muscle relaxation of the surrounding musculature.

Considering the statistical analysis adopted, should not the authors adjust the alfa based on the number of comparisons?

Response: Thank you. Based on your suggestion, we have adjusted the alfa according to the number of comparisons (2= symptomatics vs contralateral asymptomatics and symptomatics vs healthy controls), $p \leq 0.025$. This has been explicated during the data analysis section, and in Table 3, and all the manuscript has been adapted to this new consideration.

I think that the ICC, SEM, and MDC results should be in a table.

Response: We have incorporated ICC and MDC to Table 2.

Discussion

Considered the affirmation that “this was the first study to compared SUR between symptomatic and asymptomatic subjects” Actually, it was the first study to compared SUR, using the method described. Many previous studies used 3d analysis.

The authors should revise the manuscript: Haik MN, Albuquerque-Sendín F, Camargo PR. Reliability and minimal detectable change of 3-dimensional scapular orientation in individuals with and without shoulder impingement. *J Orthop Sports Phys Ther.* May 2014;44(5):341-349, which report ICC, SEM and MDC values for the upward rotation with an 3D-analysis.

Response: Thanks. Following your recommendation, we have added the following statement: “...using accessible and low-cost tools.” Regarding the study of Haik et al, 2014, we recognize its value. However, as they used a different method to assess scapular movement, we believe that their findings are not comparable with those attained in our study.

The ultrasound clinical assessment was not described in the methods section.

Response: This has been included in the methods section, inclusion criteria number 2, now it appears as follows: “(i) history of significant shoulder trauma, such as fracture or ultrasonography-clinically suspected full thickness cuff tear, following the classification of Wiener and Seitz, 1993;”.

VERSION 2 – REVIEW

REVIEWER	mark Timmons Marshall University Huntington, WV USA
REVIEW RETURNED	17-Sep-2018

GENERAL COMMENTS	The authors still do not provide a testable hypothesis. I also am not sure the authors use the correct statistical procedure to describe their data. It appears to me that some kind of ANOVA is called for.
-------------------------	--

REVIEWER	Dayana Patricia Rosa Universidad Industrial de Santander, Bucaramanga, Colombia
REVIEW RETURNED	03-Sep-2018

GENERAL COMMENTS	<p>Minor Recommendations: Authors have replied to all comments point by point. The paper addresses an important topic and should be considered for publication. However, I still have some suggestions for the authors.</p> <p>Methods:</p> <ol style="list-style-type: none">1) The authors justified why they adopted the supine position during the PM length measurement. However, they adopted the PMI from Borstad and Ludewig study (measured in standing) to classify the muscle shortening. Do not you think that it can be a bias? Please discuss this point and maybe add a limitation of the study.2) Revise the Data analysis section, please. In the manuscript is not reported about the adjusted $p < 0.025$. I think that something is lacking at the end of the first paragraph of this section. ("Based on.....")3) The ICC abbreviation is twice defined in the second paragraph of the data analysis section. <p>Results</p> <ol style="list-style-type: none">1) Please check all the manuscript to standardize the use of the abbreviations: SUR, PMI, and LSI. Sometimes, the authors use the abbreviations and sometimes not.2) Please revise the Tables 1 and 3 and define the abbreviations used in the table as you did in Table 2.3) I think the correct name of the pectoral minor is pectoralis minor. Check out this information, please.
-------------------------	---

REVIEWER	Joe Nolan Northern Kentucky University
REVIEW RETURNED	27-Nov-2018

GENERAL COMMENTS	I was asked to review this paper specifically for its statistical methods. In particular I was asked about the concern of another reviewer, who recommended that the data be analyzed using ANOVA. Based on my review, I concur with that reviewer. Ideally, ANOVA, and when applicable Repeated Measures ANOVA, should be used to analyze these data. Additional details of my review are found in the attached file. My recommendation for
-------------------------	--

major revision is based on the need for more appropriate statistical methods to be implemented.

I am performing only a statistical review of this paper. I have the following comments regarding the data analysis / methods beginning on page 10:

- The method of Kruskal-Wallis test followed by independent or paired T-tests does not seem to be appropriate. KW assumes independent samples and so would never be appropriate for pairing. Further, why is KW being used at all? It may make sense if there are substantial deviations from normality conditions. Be aware that Kolmogorov Smirnov will magnify nonnormality more than necessary (ANOVA is quite robust to mild violations of normality).
- The recommended planned approaches here would be the independent two-sample T-test for single outcomes (or ANOVA if 3 or more groups) and Repeated Measures Analysis of Variance (RM-ANOVA) for outcomes where repeated measures on the same subject exist, such as where measurements of the same thing are taken at three different positions.
- Further, and this is actually probably more important: following either ANOVA or KW by t-tests unadjusted for multiplicity is not appropriate. An adjustment must be made when doing numerous tests/comparisons. Because this study seems to consider many outcome variables, an overall Bonferroni adjustment for the number of outcome variables is appropriate. Within any ANOVA, pairwise comparisons should be conducted using the Tukey (or other similar adjustment). In no case should one revert from an ANOVA to unadjusted T-tests. Table 2 seems to begin the main section of results:
 - Table 2: I commend the authors for attempting to use confidence intervals here (this should be expressly stated in the table caption). Based on the methods section I suspect this is not done correctly. Should be based in the following:
 - o For SUR, use repeated measures ANOVA to obtain interval estimates of SUR for each of the three groups at each angle.
 - o For LSI and PMI, use a standard ANOVA to obtain interval estimates for each.
 - o ICC and MDC95 add very little to this table (and will confuse readers). They would be better positioned as a generic footnote.
 - Table 3: The idea and structure of the table are reasonable but I suspect that only basic T-tests were used to produce these and that is not correct methodology. Use ANOVA models as described above, with Tukey pairwise comparisons, to produce these results.
 - o KW test is not appropriate for SUR, given the repeated measures. If drastic violations of the normality assumption exist, consider Friedman test. Otherwise, even if small violations of normality exist, use RM-ANOVA (and note the limitation if necessary).
 - o PMI and LSI should likely be analyzed using ANOVA. KW would only be a fall-back if there are drastic violations of normality assumption.
 - o All CI's should come from the appropriate models, not from two-sample tests. It is also a bit unclear what is meant by the headings "Symptomatic-Asymptomatic shoulder", etc. Also not clear is where the healthy subjects come in. This requires further clarity.

	<p>Additional comments</p> <ul style="list-style-type: none"> • Page 10 Line 26: "Based on..." what? This appears to be a typo or a missing sentence. • Page 10 Lines 25-50: As you point out, intra-rater reliability, etc. is not a purpose of the study. It is confusing for the reader, and not particularly useful. If included, I recommend that the authors minimize its discussion so that focus is maintained on the primary analysis (e.g. in this section there is greater devotion to this than there is to statistical methods being used for the main analysis. Then on page 11 you state several numbers, which are also then included in Table 2. They should be removed from Table 2, and simplified in any reporting so as not to draw attention away from main results. I also find the idea (page 1, line 28) that this is a strength of the study to be somewhat ludicrous. It is nice to use valid/reliable measures, sure. But that doesn't necessarily make them better than other measures that have not gone through such a process, nor does it at all impact the validity of statistical methods being implemented. As you also point out on page 1, "The minimal clinically importance difference for SUR is unknown, thus we cannot make a conclusion to whether the differences found in this study mean a clinical importance or not." This fact has far greater impact than reliability measures. • Page 11, Table 1: P-values in this table are inappropriate – please remove them. This table represents a comparison of the sample demographics. There is no inference to a large population here, nor should there be. Additionally, CI's are also not appropriate. Include the sample standard deviation instead. • Pages 12, 13 – perhaps this will happen in editing, but the formatting of these tables is very poor. Perhaps put the entire confidence interval on the second line, rather than overlapping lines, if needed. • Page 15 – there is a lack of clarity regarding the numbers. Are they means and standard deviations? This should be made clear. • Page 17 – I assume "1, 15" is a typo (extra space) • Page 18 – I assume "founders" is a typo and should be "funders"
--	--

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 2

Reviewer Name: Dayana Patricia Rosa

Institution and Country: Universidad Industrial de Santander, Bucaramanga, Colombia

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Minor Recommendations: Authors have replied to all comments point by point. The paper addresses an important topic and should be considered for publication. However, I still have some suggestions for the authors.

Response: Thank you very much for your time and effort in reviewing this manuscript.

Methods:

1) The authors justified why they adopted the supine position during the PM length measurement. However, they adopted the PMI from Borstad and Ludewig study (measured in standing) to classify the muscle shortening. Do not you think that it can be a bias? Please discuss this point and maybe add a limitation of the study.

Response: Thank you. In the method section we were citing the study of Borstad et al [12] just to make reference that we had adopted the Pectoralis Minor Index (PMI) from their study. The supine position was adopted from the study of McKenzie et al. [29]. However, we can understand that this could mean bias in some way. Thus, we have added more information at this point, as follows: "Pectoralis minor index values less than 7.65 have been identified as a shortened pectoralis minor, measured in standing position[12]". In addition, we have been very carefully avoiding any comparison between this cut-off point and the results attained in this study.

2) Revise the Data analysis section, please. In the manuscript is not reported about the adjusted $p < 0.025$. I think that something is lacking at the end of the first paragraph of this section. ("Based on.....")

Response: Thank you. It is true that this part was missed in the previous version, we are sorry for that. However, based on the editor and other reviewers' suggestions, the statistical analysis has been rebuilt. Now the first paragraph of this section appears as follows: "The Statistical Package for the Social Sciences (version 23.0 for Mac; SPSS Inc. Chicago, IL) was used to analyse the collected data. Normality for all variables was explored using the Kolmogorov Smirnov test for the group of participants with shoulder pain (affected and non-affected), and for the control subjects. Two different analysis strategies were carried out: first, to determine differences in SUR at different degrees of abduction, a repeated measures ANOVA was developed in every group. For this analysis, F statistic was adjusted in case of non-sphericity (tested by Mauchly's test), with the Greenhouse-Geisser correction. Second, to determine between-groups differences for all the outcome measurements, one-way ANOVA test was calculated with Bonferroni and Tukey post-hoc estimations. A p-value less than 0.05 was considered statistically significant." Table 2 and Table 3 have changed. Even though the results have changed only a bit, all the manuscript has been adapted to the new findings.

3) The ICC abbreviation is twice defined in the second paragraph of the data analysis section.

Response: Thank you. This has been changed.

Results

1) Please check all the manuscript to standardize the use of the abbreviations: SUR, PMI, and LSI. Sometimes, the authors use the abbreviations and sometimes not.

Response: Thank you. We have revised the entire manuscript in order to identify inconsistencies. Now we believe the article has improved. On the other hand we would like to point out that in some parts of this manuscript we are describing, e.g., PMI, as a specific measurement for pectoralis minor length while in other parts we are describing pectoralis minor length in general.

2) Please revise the Tables 1 and 3 and define the abbreviations used in the table as you did in Table 2.

Response: This has been done.

3) I think the correct name of the pectoral minor is pectoralis minor. Check out this information, please.

Response: This was a typo. We have revised carefully the entire manuscript in order to avoid mistakes.

Reviewer: 1

Reviewer Name: mark timmons

Institution and Country: Marshall University, Huntington, WV USA

Please state any competing interests or state 'None declared': none

Response: Thank you very much for your time and effort in reviewing this manuscript.

Please leave your comments for the authors below
The authors still do not provide a testable hypothesis.

Response: The following statement has been incorporated to the end of the introduction: "The null hypothesis (H_0) was that there are no differences in the groups in these three different tests. The alternative hypothesis (H_a) was that there is an increased SUR in painful shoulder when comparing with contralateral and control shoulder, as well as a decreased both pectoralis minor and levator scapulae length in painful shoulder." In addition, we have included the following information at the end of the first paragraph of the discussion section: "Specifically, a decreased SUR in symptomatic shoulder compared to contralateral asymptomatic shoulder at 45 degrees, was achieved. When comparing symptomatic and control participants, an increased SUR at all positions (45, 90 and 135 degrees) was found in favour of the symptomatic shoulders. Hence, our hypothesis was only partially confirmed. Regarding PMI and LSI, there were no significant differences in the groups, thus, our hypothesis was not confirmed."

I also am not sure the authors use the correct statistical procedure to describe their data. It appears to me that some kind of ANOVA is called for.

Response: Thank you. We agree with your suggestion. Based on that and another reviewers' suggestion, we have rebuilt the statistical analysis. We believe that this section of the manuscript has greatly improved. All the tables have been changed and all the manuscript has been adapted to the new findings.

Reviewer: 3

Reviewer Name: Joe Nolan

Institution and Country: Northern Kentucky University

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

I was asked to review this paper specifically for its statistical methods. In particular I was asked about the concern of another reviewer, who recommended that the data be analyzed using ANOVA. Based on my review, I concur with that reviewer. Ideally, ANOVA, and when applicable Repeated Measures ANOVA, should be used to analyze these data. Additional details of my review are found in the attached file. My recommendation for major revision is based on the need for more appropriate statistical methods to be implemented.

Response: Thank you very much for your time and effort in reviewing this paper. We have addressed all your concerns and now we think that overall the paper is more robust and, specifically, the data analysis section.

I am performing only a statistical review of this paper. I have the following comments regarding the data analysis / methods beginning on page 10:

- The method of Kruskal-Wallis test followed by independent or paired T-tests does not seem to be appropriate. KW assumes independent samples and so would never be appropriate for pairing. Further, why is KW being used at all? It may make sense if there are substantial

deviations from normality conditions. Be aware that Kolmogorov Smirnov will magnify non-normality more than necessary (ANOVA is quite robust to mild violations of normality).

- The recommended planned approaches here would be the independent two-sample T-test for single outcomes (or ANOVA if 3 or more groups) and Repeated Measures Analysis of Variance (RM-ANOVA) for outcomes where repeated measures on the same subject exist, such as where measurements of the same thing are taken at three different positions.
- Further, and this is actually probably more important: following either ANOVA or KW by t-tests unadjusted for multiplicity is not appropriate. An adjustment must be made when doing numerous tests/comparisons. Because this study seems to consider many outcome variables, an overall Bonferroni adjustment for the number of outcome variables is appropriate. Within any ANOVA, pairwise comparisons should be conducted using the Tukey (or other similar adjustment). In no case should one revert from an ANOVA to unadjusted T-tests.

Response: Thank you. We totally agree. Thus, we have conducted a new analysis based on your recommendations. Although it was not the purpose of this study and based on your suggestion we have included in the results section a RM-ANOVA for scapular upward rotation at different angles for every single group. This has not been included in Table 2 in order to make reading easier. Thus, it has been included in the text before Table 2. Moreover, we have adjusted ANOVA using Bonferroni adjustment and pairwise comparisons have been conducted using Tukey test. Now Table 2 appears as follows:

	Symptomatic shoulder n=54	Asymptomatic shoulder n=54	Healthy controls n=40	F	p
SUR					
45° of GH abduction	4.55 (3.79 to 5.32)	5.71 (4.82 to 6.60)	2.55 (1.81 to 3.29)	F(2,145)=14.14	<0.001*
90° of GH abduction	20.75 (18.81 to 22.69)	21.42 (19.88 to 22.96)	16.77 (15.49 to 18.04)	F(2,145)=8.08	<0.001*
135° of GH abduction	45.18 (42.76 to 47.59)	44.16 (42.20 to 46.12)	36.22 (34.34 to 38.09)	F(2,145)=18.64	<0.001*
LSI	7.81 (7.42 to 8.20)	7.81 (7.53 to 8.30)	7.76 (7.42 to 8.11)	F(2,145)=0.02	0.978
PMI	10.52 (10.27 to 10.76)	10.86 (10.26 to 11.46)	10.07 (9.73 to 10.42)	F(2,145)=2.97	0.054

Table 2: Mean values of pectoralis minor index (PMI), levator scapulae index (LSI), and scapular upward rotation expressed in degrees (SUR) in different groups; F: One-factor ANOVA for differences in symptomatic, asymptomatic and healthy controls.

*: statistically significant.

Table 2 seems to begin the main section of results:

- Table 2: I commend the authors for attempting to use confidence intervals here (this should be expressly stated in the table caption). Based on the methods section I suspect this is not done correctly. Should be based in the following:

- o For SUR, use repeated measures ANOVA to obtain interval estimates of SUR for each of the three groups at each angle.

- o For LSI and PMI, use a standard ANOVA to obtain interval estimates for each

- o ICC and MDC95 add very little to this table (and will confuse readers). They would be

better positioned as a generic footnote.

Response: This information has been inserted in Table 2 (see above) and in the text just before Table 2, as follows: "Mean values of scapular upward rotation (expressed in degrees), levator scapulae index (LSI) and pectoralis minor index (PMI) for all the groups are presented in Table 2. There were statistically significant differences in SUR when comparing the three groups, while no differences were found for the rest of the outcome measurements (LSI and PMI) (see Table 2). Furthermore, there was an increase in SUR from 45 to 90 and 135 degrees of shoulder abduction for all the groups, analysed by repeated measures ANOVA, with the following results:
 Symptomatic shoulder: $F(1.51, 80.05) = 1009.22; p < 0.001$
 Asymptomatic shoulder: $F(1.46, 77.37) = 1356.57; p < 0.001$
 Healthy controls: $F(1.46, 56.89) = 1196.18; p < 0.001$

• Table 3: The idea and structure of the table are reasonable but I suspect that only basic T-tests were used to produce these and that is not correct methodology. Use ANOVA models as described above, with Tukey pairwise comparisons, to produce these results.

Results: This has been done as explained aforementioned. Now Table 3 appears as follows:

	Symptomatic vs Asymptomatic shoulder differences (95%CI)	p	Symptomatic vs Control shoulder differences (95%CI)	p
SUR				
At 45° GH abduction	-1,15 (-2,46 to -0,15)	0.09	2,01 (0,59 to 3,42)	0.003*
At 90° GH abduction	-0,67 (-3,35 to 2)	0.82	3,98 (1,08 to 6,88)	0.004*
At 135° GH abduction	1,02 (-2,41 to 4,45)	0.76	8,96 (5,24 to 12,6)	<0.001*
PMI	-0,34% (-1,04 to 0,36%)	0.49	0,45% (-0,32 to 1,21%)	0.351
LSI	0,00% (-0,55 to 0,55%)	1	0.05% (-0,55 to 0,64%)	0.98

Table 3: Between- differences group (Bonferroni and Tukey multiple comparisons)

CI: confidence interval; SUR: scapular upward rotation; GH: glenohumeral; PMI: pectoralis minor index;

LSI: levator scapulae index

*: statistically significant ($p < 0.05$)

o KW test is not appropriate for SUR, given the repeated measures. If drastic violations of the normality assumption exist, consider Friedman test. Otherwise, even if small violations of normality exist, use RM-ANOVA (and note the limitation if necessary).

Response: This has been done as aforementioned.

o PMI and LSI should likely be analyzed using ANOVA. KW would only be a fall-back if there are drastic violations of normality assumption.

Response: PMI and LSI measurements have been analysed using ANOVA, as well as SUR measurements, during the comparisons in all the groups, and, as aforementioned, Bonferroni adjustments and Tukey pairwise comparisons have been carried out.

o All CI's should come from the appropriate models, not from two-sample tests. It is also a bit unclear what is meant by the headings "Symptomatic-Asymptomatic shoulder", etc. Also not clear is where the healthy subjects come in. This requires further clarity.

Response: All the confidence intervals (CIs) have been reported according to the findings attained when ANOVA and, consequently, Bonferroni adjustments and Tukey pairwise comparisons, have been carried out. Regarding the headings "Symptomatic-Asymptomatic shoulder", we have changed a bit this statement, and now it appears: "Symptomatic vs Asymptomatic shoulder", etc. This has been explicated across the entire manuscript, e.g., hypothesis of the study, and data analysis section. In one hand, a one-way ANOVA was carried out to determine differences in all the groups (Table 2). On the other hand, pairwise comparisons were carried out to determine differences between:

- Symptomatic versus asymptomatic shoulder
- Symptomatic versus healthy (non-painful) shoulder

These comparisons are supported by the rationale behind the whole manuscript. The report of the differences between asymptomatic and healthy controls was not informed as this was not the purpose of this study.

Additional comments

- Page 10 Line 26: "Based on..." what? This appears to be a typo or a missing sentence.

Response: Thank you. This was a mistake. All the data analysis section has been modified as aforementioned.

- Page 10 Lines 25-50: As you point out, intra-rater reliability, etc. is not a purpose of the study. It is confusing for the reader, and not particularly useful. If included, I recommend that the authors minimize its discussion so that focus is maintained on the primary analysis (e.g. in this section there is greater devotion to this than there is to statistical methods being used for the main analysis. Then on page 11 you state several numbers, which are also then included in Table 2. They should be removed from Table 2, and simplified in any reporting so as not to draw attention away from main results. I also find the idea (page 1, line 28) that this is a strength of the study to be somewhat ludicrous. It is nice to use valid/reliable measures, sure. But that doesn't necessarily make them better than other measures that have not gone through such a process, nor does it at all impact the validity of statistical methods being implemented. As you also point out on page 1, "The minimal clinically importance difference for SUR is unknown, thus we cannot make a conclusion to whether the differences found in this study mean a clinical importance or not." This fact has far greater impact than reliability measures.

Response: Thank you. First, we have removed ICC values from Table 2, as suggested. They have been included into the text after Table 2. Second, we have deleted from the strengths the statement regarding the excellent ICC values achieved in the study.

- Page 11, Table 1: P-values in this table are inappropriate – please remove them. This table represents a comparison of the sample demographics. There is no inference to a large population here, nor should there be. Additionally, CI's are also not appropriate. Include the sample standard deviation instead.

Response: Thank you. P-values have been removed from Table 1. CI's have been replaced by SDs.

- Pages 12, 13 – perhaps this will happen in editing, but the formatting of these tables is very poor. Perhaps put the entire confidence interval on the second line, rather than overlapping lines, if needed.

Response: This has been done.

- Page 15 – there is a lack of clarity regarding the numbers. Are they means and standard deviations? This should be made clear.

Response: Thank you. All of these results were expressed in terms of means and standard deviations. This has been explicated in the text.

- Page 17 – I assume “1, 15” is a typo (extra space)

Response: This was a typo and it has been corrected.

- Page 18 – I assume “founders” is a typo and should be “funders”

Response: This has been solved.

VERSION 3 - REVIEW

REVIEWER	Dayana Patricia Rosa Brazil
REVIEW RETURNED	20-Jan-2019

GENERAL COMMENTS	<p>Authors have replied to all comments point by point. The paper addresses an important topic and should be considered for publication. I have only quick suggestions for the authors, considering the alterations done.</p> <p>It would be interesting to add the evaluated shoulder to Table 1.</p> <p>I recommend describing the ANOVA results showed in page 43 as a text and not as a topic.</p> <p>What are the F values in between parentheses at the beginning of page 43 and in Table 2?</p> <p>I suggest reporting the ICC values as a table.</p>
-------------------------	--

REVIEWER	Joe Nolan Northern Kentucky University
REVIEW RETURNED	03-Jan-2019

GENERAL COMMENTS	<p>I am comfortable that mostly proper statistical methods are now employed in the manuscript; additionally I applaud the fact that a statistician author was added. I saw only two remaining minor points that can quickly be resolved:</p> <p>1. The suggestion of a Bonferroni adjustment was not intended to be for pairwise comparisons - those should be Tukey. Rather, Bonferroni was intended at the global level. You have at least 3 response variables: SUR, PMI, and LSI. If I haven't missed any, that means your initial alpha level should be $0.05 / 3 = 0.0167$ (a Bonferroni adjustment). P-values for F-tests, etc. would be compared to 0.0167 rather than 0.05. Then within an ANOVA you would also apply Tukey while doing pairwise comparisons.</p> <p>2. I reiterate that the component on inter-rater reliability is of value to the paper. I would leave it out completely. It is a topic that certainly should not comprise half of the methods section. One sentence (without formulae) at most. Anything more draws the reader's focus away from your main study and results.</p>
-------------------------	---

VERSION 3 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 3

Reviewer Name: Joe Nolan

Institution and Country: Northern Kentucky University

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

I am comfortable that mostly proper statistical methods are now employed in the manuscript; additionally I applaud the fact that a statistician author was added. I saw only two remaining minor points that can quickly be resolved:

Response: Thank you very much for your time and effort in reviewing this paper.

1. The suggestion of a Bonferroni adjustment was not intended to be for pairwise comparisons - those should be Tukey. Rather, Bonferroni was intended at the global level. You have at least 3 response variables: SUR, PMI, and LSI. If I haven't missed any, that means your initial alpha level should be $0.05 / 3 = 0.0167$ (a Bonferroni adjustment). P-values for F-tests, etc. would be compared to 0.0167 rather than 0.05.

Then within an ANOVA you would also apply Tukey while doing pairwise comparisons.

Response: Thank you very much. We agree with that. The suggested changes have been done.

2. I reiterate that the component on inter-rater reliability is of value to the paper. I would leave it out completely. It is a topic that certainly should not comprise half of the methods section. One sentence (without formulae) at most. Anything more draws the reader's focus away from your main study and results.

Response: Thank you. This has been done. We have eliminated all the information regarding ICC from the results section. Likewise, we have reduced the long description on intra-rater reliability from the data analysis section, as follows: "The intraclass correlation coefficient was greater than 0.90 for all the tests, which means an excellent reliability[33], except for LSI (0.87). The MDC95 was as follows: SUR45°= 0.91; SUR90°= 1.55; SUR135°= 2.83; PMI= 0.80; LSI= 1.08."

Reviewer: 2

Reviewer Name: Dayana Patricia Rosa

Institution and Country: Brazil

Please state any competing interests or state 'None declared':None declared

Please leave your comments for the authors below

Authors have replied to all comments point by point. The paper addresses an important topic and should be considered for publication. I have only quick suggestions for the authors, considering the alterations done.

Response: Thank you again for your time and effort in reviewing this paper.

It would be interesting to add the evaluated shoulder to Table 1.

Response: If you are referring to the outcome variables of symptomatic shoulders (SUR, PMI AND LSI), these values are shown in Table 2. If you are referring to which shoulder was evaluated in patients, this statement can be found in the method section: "A sample of seventy-three patients with chronic SAPS in their dominant arm...". Additionally, we have included in Table 1 information regarding which shoulder is being referenced.

I recommend describing the ANOVA results showed in page 43 as a text and not as a topic.

Response: This has been done.

What are the F values in between parentheses at the beginning of page 43 and in Table 2?

Response: These F values are coming from different analysis.

Analysis 1

The following statement/findings are coming from the analysis about how every group (symptomatic, asymptomatic and healthy groups) changed their SUR values across the different shoulder elevation degrees, as suggested in the below paragraph:

“Furthermore, there was an increase in SUR from 45 to 90 and 135 degrees of shoulder abduction for all the groups, analysed by repeated measures ANOVA, with the following results: (i) symptomatic shoulder: $F(1,51; 80.05) = 1009.22$; $p < 0.001$; (ii) asymptomatic shoulder: $F(1,46; 77.37) = 1356.57$; $p < 0.001$; (iii) healthy controls: $F(1,46; 56.89) = 1196.18$; $p < 0.001$ ”

In this context, F values are reported, e.g., 1009.22, for symptomatic group, with 1 degree of freedom. The value “77.37” makes reference to Greenhouse-Geisser correction, as cited in the method section.

Analysis 2

Regarding Table 2, the F values show differences in every outcome across the different groups, e.g., $F(2,145) = 14.14$, where F value = 14.14, with 2 degrees of freedom.

I suggest reporting the ICC values as a table.

Response: As one of the reviewers (statistician) has suggested in a previous and the present review, we have considerably reduced the information regarding the intra-rater reliability and, thus, ICC values. This has been mentioned in the data analysis section, as follows: “The intraclass correlation coefficient was greater than 0.90 for all the tests, which means an excellent reliability [33], except for LSI (0.87). The MDC95 was as follows: $SUR_{45^\circ} = 0.91$; $SUR_{90^\circ} = 1.55$; $SUR_{135^\circ} = 2.83$; $PMI = 0.80$; $LSI = 1.08$.”

VERSION 4 - REVIEW

REVIEWER	Joe Nolan Northern Kentucky University
REVIEW RETURNED	28-Feb-2019

GENERAL COMMENTS	The implementation of statistics in this paper has greatly improved. I do note that there are still a number of typos in the paper (to which the list below should not be assumed comprehensive):
-------------------------	---

	<p>Page 10 line 49: "was used to analyses the collected data" is poor phrasing. (to analyze?)</p> <p>Table 1 caption: Should specify SD rather than 95%CI.</p> <p>Table 3: Decimal points were used in all other tables, why are comma's being used here? Such formatting ought to be consistent.</p>
--	---