PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	ULTRASONIC STRAIN ELASTOGRAPHY FOR DETECTING ABNORMALITIES IN THE SUPRASPINATUS TENDON: AN INTRA- AND INTER-RATER RELIABILITY STUDY
AUTHORS	Brage, K; Hjarbaek, John; Kjaer, Per; Ingwersen, Kim G.; Juul-Kristensen, Birgit

VERSION 1 - REVIEW

REVIEWER	Takayuki Muraki Department of Physical Medicine and Rehabilitation, Tohoku University Graduate School of Medicine, Japan
REVIEW RETURNED	21-Dec-2018

GENERAL COMMENTS	General comments The authors investigate the reliability of the measurement of tissue elasticity by using strain-elastography. Although some previous studies (Seo et al, 2015; Muraki et al, 2015) demonstrated the reliability, this study have the strength that both patients with supraspinatus tendinopathy and healthy subjects are enrolled. However, I have several concerns that need to be addressed before considering publication. The authors should clarify the points listed below. In addition, there are a number of typos (e.g., P4 line 13: and absence from work,[4].→and absence from work [4].). Furthermore, there are the use of incorrect grammar and verb tenses, so the manuscript needs a thorough linguistic revision by a native English speaker or a professional editor. Specific comments Introduction (1) Page 4, line 34 Strain elastography (SEL) is a relatively new and not yet well-established method, which may assist in early diagnosis, prediction and monitoring of progress in tendon healing,[7]. →I cannot understand why SEL assist in early diagnosis. This seems to be based more on speculation. Please clearly describe the reason on the basis of evidence. (2) Page 5, line 27 the validity of SEL in the supraspinatus tendon seems promising,[13-16]. →The term of ”validity” may be somewhat misleading. I suggest that the authors replace “the validity” by “the availability”.

	(3) Page 5, line 48 However, this study did not include a healthy control group and used only color quantification,[14]. →"this study did not include a healthy control group" > I don't think that this is a limitation in the previous study. If the authors consider that this is one of the limitations, please include more explanation on the reason. (4) Page 5, line 55 Limitations of this study were; not including a group with pathology, not defining ROI and only using one type of reference tissue (acoustic coupler),[16]. →"only using one type of reference tissue (acoustic coupler)" > Why do the authors consider the assessment of the reliability using more than one reference to be necessary. In particular, the living tissue (i.e., deltoid muscle) is not suitable as a reference material because the elasticity of the deltoid muscle is subject to change. The authors need to clatify the objective of using the deltoid muscle as a reference material. Materials and methods (5) Page 9, line 9 Furthermore, at least one shoulder had to be diagnosed with tendinosis based on MRI (≥ grade 1),[21]. →There appears to be a lack of explanation on the diagnostic criteria for supraspinatus tendinosis. I think this is very important information for the inclusion of subjects. The patients shuold be diagnosis with supraspinatus tendinosis on the basis of the gold standard criteria. Please add description of the diagnostic criteria in more detail. (6) Page 9, line 16 By convenience 50% of participants had their dominant arm scanned with strain elastography, while the remaining 50% had their non-dominant arm scanned. →I cannot understand what meaning of this sentence. Was this procedure conducted for adjusting a confounding factor (hand dominance)? If so, was this allocation randomized? (7) Page 11, line 7 The gel pad was used as a more homogeneous reference tissue →Please provide more information to understand what the gel pad is in (size, material, and commercialization…). In addition, please clarify the elastic modulus. (8) Page 12, line 55 Hereafter, phase three (the actual study phase, n=40) was initiated based on the final protocol, as described above. →Please provide detail regarding the sample size calculation to justify a sample of n=40. You need to add description of the statistical power. Results (9) Page 17, line 19 The same pattern, for all measurements, was seen for each of the three measured areas of the supraspinatus tendon (not shown in tables). →I cannot understand what meaning of this sentence. What is the same pattern? Please explain it in more detail.

	Discussion (10) Page 20, line 7 The reliability of the results from SEL was 'excellent' when using the raw data and the deltoid muscle as reference tissue. When using a gel pad as reference the reliability of the results was 'good'. Furthermore, the relative MDC (% of mean) was smallest for RAW and largest for GEL. →The authors give no explanation why the ICC value of gel pad was lower than that of the raw data and the deltoid muscle. In theory, the reliability when using a gel pad seems to be higher because the elasticity of the gel pad is homogeneous and unchanged. It is difficult for me to interpret this result. Although the authors mentioned "The present study found the lowest reliability (but still graded 'good') when using GEL which may be caused by difficulties due to lower image quality, because of increased depth (Page 22, line 34)", there is no evidence to support this. In addition, previous studies (Muraki et al, 2015; Yamamoto et al, 2016) demonstrated excellent reliability by using an acoustic coupler that similar to the gel pad. Please thoroughly discuss why the reliability of gel pad was lower than that of the deltoid muscle in the Disccusion section. (11) Page 22, line 4 The present study found a high MDC for GEL (large measurement error) but a good reliability (ICC 0.70-0.73), in line with previous studies. →I think that MDC for GEL (Intra-rater, 58.82%; Inter-rater, 62.63%) is too much higher, which means large measurement error. It is doubtful whether this measurement using GEL is clinically available. On the other hand, the present study showed that ICC values (Intra-rater, 0.73; Inter-rater, 0.70) is higher, which means good reliability. How can the authors explain the discrepancy between the results for MDC and ICC? →Previous studies (Muraki et al, 2015; Yamamoto et al, 2016) demonstrated excellent reliability by using an acoustic coupler. However, the reliability using the gel pad in the present study was lower compared to the results of previous studies. You need to describe what is contribute to this discrepancy between the results of the previous study and present study. (12) Page 23, line 36 The reason may be due to the low presence of lesions in the lateral part, corresponding to only 10 % of the participants presenting with lesions in the lateral 1/3 of the tendon… →Is this fact supported clearly enough? If so, you should provide a reference to justify the fact. Table (13) Page 15 (Table 1) "Gender (n (females (%))" > This is confusing description. Please indicate the number of men and women. (14) Page 16 (Table 2) "0.00*" > The asterisk means significant difference between Rater 1 and Rater 2. Hence, the authors need to change the asterisk into other symbol (** or †) and add the description.

REVIEWER	Gordana Ivanac Department of Diagnostic and Interventional Radiology, Clinical Hospital Dubrava, Zagreb, Croatia
REVIEW RETURNED	24-Jan-2019

GENERAL COMMENTS	Previously published study compared only strain elastography and MRI in patients with shoulder pathology and in another study the same methods were tested in healthy individuals. So far, this is the first reliability study that includes both patients and healthy controls. The aim of the present study was to assess the inter and intra-rater reliability of SEL within the tendon of m. supraspinatus by using different reference tissues and quantification methods, and it was clearly presented. Validated guidelines for reporting reliability were used, and study was structured in 3 phases to avoid clinician dependency and systemic bias. Collectively, SEL was shown as reliable and reproducible method for recognizing the pathology of supraspinatus tendon. Among healthy individuals authors did not state the difference between dominant and non-dominant hand, if it is any.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Takayuki Muraki

Institution and Country: Department of Physical Medicine and Rehabilitation, Tohoku University Graduate School of Medicine, Japan

Question A): Please leave your comments for the authors below.

General comments: The authors investigate the reliability of the measurement of tissue elasticity by using strain-elastography. Although some previous studies (Seo et al, 2015; Muraki et al, 2015) demonstrated the reliability, this study have the strength that both patients with supraspinatus tendinopathy and healthy subjects are enrolled. However, I have several concerns that need to be addressed before considering publication. The authors should clarify the points listed below. In addition, there are a number of typos (e.g., P4 line 5: and absence from work,[4].→and absence from work [4].

Answer: Thank you for your general comment on the strengths of the trial and for pointing out the problems with the formatting of references.

Action: All references have been updated to the correct format

Question B): Furthermore, there are the use of incorrect grammar and verb tenses, so the manuscript needs a thorough linguistic revision by a native English speaker or a professional editor.

Answer: Thank you for this suggestion. The manuscript has now gone through linguistic revision by an English native speaker.

Action: Different sentences have been changed accordingly (only in the clean copy).

Specific comments

Introduction

Question 1): Page 4, line 14-15, Strain elastography (SEL) is a relatively new and not yet well-established method, which may assist in early diagnosis, prediction and monitoring of progress in tendon healing,[7].

→I cannot understand why SEL assist in early diagnosis. This seems to be based more on speculation. Please clearly describe the reason on the basis of evidence.

Answer: We agree, that early diagnosis is not an optimal word.

Action: The word early has now been removed from the sentence.

Question 2): Page 5, line 12. the validity of SEL in the supraspinatus tendon seems promising,[13-16].

→The term of ”validity” may be somewhat misleading. I suggest that the authors replace “the validity” by “the availability”.

Answer: We agree that the term ‘validity’ is a broad term and too vaguely defined. However, as the sentence referes to the validity and not the general availability of SEL, we have more clearly defined the type of validity we describe.

Action: The following has now been inserted: ’the concurrent validity of SEL in the supraspinatus tendon seems promising‘,[13-15].

Question 3): Page 5, line 23-24. However, this study did not include a healthy control group and used only color quantification,[14].

→”this study did not include a healthy control group” > I don’t think that this is a limitation in the previous study. If the authors consider that this is one of the limitations, please include more explanation on the reason.

Answer: We agree and have specified the meaning of this.

Action: The sentence has now been altered to:… ‘However, this study did not use a healthy control group which is recommended for reliability- and validity studies of today, in order to achieve large variation) 2. Furthermore, the study used only color quantification’.

Question 4): Page 5, line 23. Limitations of this study were; not including a group with pathology, not defining ROI and only using one type of reference tissue (acoustic coupler),[16].

→”only using one type of reference tissue (acoustic coupler)” > Why do the authors consider the assessment of the reliability using more than one reference to be necessary. In particular, the living tissue (i.e., deltoid muscle) is not suitable as a reference material because the elasticity of the deltoid muscle is subject to change. The authors need to clatify the objective of using the deltoid muscle as a reference material.

Answer: Standardised and consensus procedures for conducting elastography in the supraspinatus tendon have not yet been established. Therefore, there is a need to investigate, which reference tissue has the highest reliability. We agree that the deltoid may not be optimal as a reference, but have included it as an exploratory method in addition to the gelpad.

Action: The following sentence has now been added: ‘Since standardised and consensus procedures for conducting elastography in the supraspinatus tendon have not yet been established, there is a need to investigate, which reference tissue has the highest reliability.

Question 5): Materials and methods. Page 9, line 3. Furthermore, at least one shoulder had to be diagnosed with tendinosis based on MRI (≥ grade 1),[21].

→There appears to be a lack of explanation on the diagnostic criteria for supraspinatus tendinosis. I think this is very important information for the inclusion of subjects. The patients shuold be diagnosis with supraspinatus tendinosis on the basis of the gold standard criteria. Please add description of the diagnostic criteria in more detail.

Answer: Thank you for this suggestion. The sentence has now been altered.

Action: 'Furthermore, at least one shoulder had to be diagnosed with tendinosis clinically (positive signs of ≥ 3 clinical tests (Hawkins-Kennedy test; Neers test; Empty Can test; Full Can test; Resisted external rotation test)) and based on MRI (≥ grade 1; corresponding to focal increase in tendon signal on protondensity weighted- and fatsuppressed T2 sequences not equal to that of fluid,[21]).'

Question 6):Page 9, line 6. By convenience 50% of participants had their dominant arm scanned with strain elastography, while the remaining 50% had their non-dominant arm scanned.

→I cannot understand what meaning of this sentence. Was this procedure conducted for adjusting a confounding factor (hand dominance)? If so, was this allocation randomized?

Answer: Thank you for this suggestion. The sentence has now been changed and the frequencies are further shown in table 1.

Action: '…For the healthy participants the choice of shoulder was matched with that of the patients.

Question 7): Page 10, line 45. The gel pad was used as a more homogeneous reference tissue
→Please provide more information to understand what the gel pad is in (size, material, and commercialization…). In addition, please clarify the elastic modulus.

Answer: Thank you for your request to elaborate on this. In our review of the sentence we discovered a mistake in the described manufacturer, why this also have been changed.

Action: For each assessment method (with or without a gel pad covering the transducer "(Sonokit (Proxon), thickness: 10 mm, length: 70 mm, elastic modulus: 226 kPa; Sonogel Vertriebs GmbH, Germany))" 3 sessions of 20 sec. were obtained.

Question 8): Page 12, line 23. Hereafter, phase three (the actual study phase, n=40) was initiated based on the final protocol, as described above.

→Please provide detail regarding the sample size calculation to justify a sample of n=40. You need to add description of the statistical power.

Answer: The study followed the protocol for diagnostic procedures in reproducibility studies 2. The phases constitute a methodological model for optimising procedures, and aim at eliminating clinician subjectivity as much as possible. The aim of the training phase is to ensure that raters have sufficient competence and experience in performing the procedures. The overall agreement phase is an extended training phase and ensures that gross systematic bias between raters is minimised, and requires at least 80% agreement between raters before proceeding to phase 3. The study phase is the final evaluation of reliability of the developed procedures 2. The number of participants recommended by Patijn, 2004 is based on a pragmatic number to ensure enough participants to obtain acceptable reliability, when the 3-phases standardized approach is used.

Action: None.

Question 9): Results. Page 17, line 7. The same pattern, for all measurements, was seen for each of the three measured areas of the supraspinatus tendon (not shown in tables).

→I cannot understand what meaning of this sentence. What is the same pattern? Please explain it in more detail.

Answer: Thank you for letting us explain this more precisely. The precented data er for the mean across the tendon includind both the midial part, the middle part and the lateral part.

Action: The sentence has been altered to: '…The same pattern, for all measurement types (RAW/DELT: excellent agreement, GEL: good agreement), was also seen for each of the three measured areas of the supraspinatus tendon (medial-, middle- and lateral part) (not shown in tables).'

Question 10): Discussion. Page 20, line 2. The reliability of the results from SEL was 'excellent' when using the raw data and the deltoid muscle as reference tissue. When using a gel pad as reference the reliability of the results was 'good'. Furthermore, the relative MDC (% of mean) was smallest for RAW and largest for GEL.

→The authors give no explanation why the ICC value of gel pad was lower than that of the raw data and the deltoid muscle. In theory, the reliability when using a gel pad seems to be higher because the elasticity of the gel pad is homogeneous and unchanged. It is difficult for me to interpret this result. Although the authors mentioned "The present study found the lowest reliability (but still graded 'good') when using GEL which may be caused by difficulties due to lower image quality, because of increased depth (Page 22, line 14)", there is no evidence to support this. In addition, previous studies (Muraki et al, 2015; Yamamoto et al, 2016) demonstrated excellent reliability by using an acoustic coupler that similar to the gel pad. Please thoroughly discuss why the reliability of gel pad was lower than that of the deltoid muscle in the Disccusion section.

Answer: You are right that the gel pad is more homogeneous. However, since we used a high frequency transducer with better detail resolution, penetration is lower 3. This means that adding an extra 10 mm gelpad, between the transducer and the gelpad, will result in lower image quality. Previous studies 4 5 have measured the ROI in the middle of the tendon, but the exact location was unclear.

The present study measured the ROI of the tendon exactly relative to the footprint in 3 parts: the lateral-, middle- and the medial part. When using the gelpad a lower quality was found as it was often difficult to locate the footprint, which thereby may have reduced the ICC. This is a limitation of the method, but only when using the gelpad. In contrast, the strengths of using the deltoid as reference instead, is that this tissue is placed close to the tendon and therefore this reference value is influenced by the same (or close to) pressure of the tendon, as opposed to the pressure on the gelpad.

Action:The following sentence (page 22, line 14) has now been changed:..The present study found the lowest reliability (but still graded 'good') when using GEL which may be caused by difficulties "locating the footprint of the tendon" due to lower image quality, because of increased depth "(through 10 mm gelpad)". Also a statistical significant intra-rater difference was seen, but as this difference was below the MDC it can be ascribed as measurement error.

Question 11): Page 22, line 1. The present study found a high MDC for GEL (large measurement error) but a good reliability (ICC 0.70-0.73), in line with previous studies.

→I think that MDC for GEL (Intra-rater, 58.82%; Inter-rater, 62.63%) is too much higher, which means large measurement error. It is doubtful whether this measurement using GEL is clinically available. On the other hand, the present study showed that ICC values (Intra-rater, 0.73; Inter-rater, 0.70) is

higher, which means good reliability. How can the authors explain the discrepancy between the results for MDC and ICC?

→Previous studies (Muraki et al, 2015; Yamamoto et al, 2016) demonstrated excellent reliability by using an acoustic coupler. However, the reliability using the gel pad in the present study was lower compared to the results of previous studies. You need to describe what is contribute to this discrepancy between the results of the previous study and present study.

Answer: Thank you for letting us explain this more preceisely. The present ICC is reported as single measures (described in statistical section), which is reported as it is of relevance for clinicians. Neither Muraki nor Yamamoto have stated whether ICC is reported as single measures or (group) average measure, but it is anticipated that it is group average measures. In the present data ICC will increase about 10% when reported as average measures, and MDC will decrease withup to 75%. This could explain some of the difference between MDC and ICC.

Action: Page 22, line 6. The following sentence has been inserted after “Unfortunately MDC was not reported”. The present ICC and MDC are reported as single measures, as it is of relevance for clinicians. None of the previous studies (Muraki, Yamamoto) have decribed whether ICC is reported as single measures or (group) average measure, however, it is anticipated that it is group average measures. The present ICC will increase about 10% and the MDC will decrease with up to 80% when reported as average measures. This could explain some of the differences in results between the present and the previous studies.’

Question 12): Page 23, line 16-17. The reason may be due to the low presence of lesions in the lateral part, corresponding to only 10 % of the participants presenting with lesions in the lateral 1/3 of the tendon… →Is this fact supported clearly enough? If so, you should provide a reference to justify the fact.

Answer: Thank you. This has now been explained more clearly.

Action: The following sentence has now been inserted: ‘The reason may be due to the current low presence of lesions in the lateral part, corresponding to only 10 % of the participants, showing lesions in the lateral 1/3 of the tendon which can lead to the ‘Kappa Paradox’,6. The kappa paradox means (in a 2x2 table) that an imbalance between presence and absence in overall agreement, and between disagreement (bias) can drastically lower the kappa, why prevalence- and bias adjusted kappa (PABAK) in dichotomous scales is recommended 7. A PABAK adjusted kappa on the present number of leasions dicotomised into; no leasion/leasion(s), will increase the kappa by up to 70%.

Question 13): Page 15 (Table 1). “Gender (n (females (%))” > This is confusing description. Please indicate the number of men and women.

Answer: This has been altered.

Action: Table 1 has been changed.

Question 14): Page 16 (Table 2). “0.00*” > The asterisk means significant difference between Rater 1 and Rater 2. Hence, the authors need to change the asterisk into other symbol (** or †) and add the description.

Answer: Thank you. Instead of using two asterisk sympols with explanations we have chosen to simplify the description af the asterisk.

Action: The table 2 description has been altered from “significant difference between first and second measurements for Rater 1 and 2” to “significant difference ($p \leq 0.05$) between measurements”.

Reviewer: 2

Reviewer Name: Gordana Ivanac

Institution and Country: Department of Diagnostic and Interventional Radiology, Clinical Hospital Dubrava, Zagreb, Croatia

Question 1):

Please leave your comments for the authors below: Previously published study compared only strain elastography and MRI in patients with shoulder pathology and in another study the same methods were tested in healthy individuals. So far, this is the first reliability study that includes both patients and healthy controls. The aim of the present study was to assess the inter and intra-rater reliability of SEL within the tendon of m. supraspinatus by using different reference tissues and quantification methods, and it was clearly presented. Validated guidelines for reporting reliability were used, and study was structured in 3 phases to avoid clinician dependency and systemic bias. Collectively, SEL was shown as reliable and reproducible method for recognizing the pathology of supraspinatus tendon.

Among healthy individuals authors did not state the difference between dominant and non-dominant hand, if it is any.

Answer: Thank you. This has now been included.

Action: The difference in dominant and non-dominant measurements between patients and healthy participants has now been inserted in table 1.

References

1. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011;64(1):96-106. doi: 10.1016/j.jclinepi.2010.03.002

2. Patijn J. Reproducibility and Validity Studies of Diagnostic Procedures in Manual/Musculoskeletal Medicine 2004.

3. Kremkau FW. Sonography 2011.

4. Muraki T, Ishikawa H, Morise S, et al. Ultrasound elastography-based assessment of the elasticity of the supraspinatus muscle and tendon during muscle contraction. Journal of shoulder and elbow surgery

5. Yamamoto Y, Yamaguchi S, Sasho T, et al. Quantitative Ultrasound Elastography With an Acoustic Coupler for Achilles Tendon Elasticity: Measurement Repeatability and Normative Values. Journal of ultrasound in medicine

6. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543-9.

7. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993;46(5):423-9.

VERSION 2 – REVIEW

REVIEWER	Takayuki Muraki Department of Physical Medicine and Rehabilitation, Tohoku University Graduate School of Medicine, Japan
REVIEW RETURNED	28-Mar-2019

GENERAL COMMENTS	General comments The manuscript has been revised well and updated to the correct format. I agreed that you replied satisfactorily to our concerns. In particular, I understood about the 3-phases standardized approach for reporting reliability. In discussion section, the authors gave explanation why the ICC value of gel pad was lower than that of the raw data and the deltoid muscle, which is caused by difficulties “locating the footprint of the tendon” and lower image quality. I think that it is satisfactory explanation. Therefore, I have no more comments to make.