



## REVIEW

# Quality of reliability studies for ultrasound evaluation of Achilles tendon: A systematic review

Elena Portaceli-Oñate<sup>a</sup>, Federico D. Puglisi<sup>b</sup>, Jacinto J. Martínez-Payá<sup>a</sup>,  
M<sup>a</sup> Elena del Baño-Aledo<sup>a</sup> and José Ríos-Díaz<sup>a,b,\*</sup>

<sup>a</sup>G.I ECOFISTEM. Facultad de Ciencias de la Salud. Universidad Católica de Murcia, Spain

<sup>b</sup>Centro Universitario de Ciencias de la Salud San Rafael-Nebrija. Madrid, Spain

### KEYWORDS

Reliability;  
Ultrasonography;  
Achilles tendon;  
QAREL

### Abstract

**Introduction:** The Achilles tendon is one of the most frequently injured tendons. Ultrasound is a non-invasive, painless and inexpensive technique which is increasingly used in clinical practice for assessing the condition of tendons. However, the reliability of the same depends on the experience and technical skill of the examiner. The aim of this study was to review and assess the methodological quality of reliability studies for ultrasound examination of the Achilles tendon.

**Material and method:** A systematic review was performed following the PRISMA guidelines. We searched MEDLINE, SPORTDiscus, Academic Search, CINAHL, E-Journals, IBECS, LILACS OVID, DIALNET, COCHRANE and PEDro electronic databases. We created a tool consisting of 9 items to assess the methodological quality of the description of ultrasound examinations. The QAREL scale was used to assess the methodological quality of the selected reliability studies.

**Results:** 227 articles were identified and, after a selection process, 13 articles were chosen for inclusion in this review. The information provided regarding the blinding of examiners and statistical analysis was insufficient in 70% and 69% of the analysed cases, respectively.

**Conclusions:** The description of the sonographic method was insufficient to ensure the reproducibility of the studies reviewed. Furthermore, the articles have methodological issues in relevant aspects, such as the representativeness of the sample, blinding of examiners and the appropriate statistical analysis of the data.

© 2016 MVclinic. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\*Corresponding author.

E-mail address: [Jrios@nebrija.es](mailto:Jrios@nebrija.es) (J. Ríos-Díaz).

**PALABRAS CLAVE**

Fiabilidad;  
Ultrasonografía;  
Tendón de Aquiles;  
QAREL

## Calidad de los estudios de fiabilidad de la exploración ecográfica en el tendón de Aquiles: Revisión sistemática

**Resumen**

**Introducción:** El tendón de Aquiles es uno de los tendones que más frecuentemente se lesiona. La ecografía, una técnica no invasiva, indolora y de bajo coste, tiene cada vez más importancia en la práctica clínica para valorar el estado del tendón; sin embargo, su fiabilidad depende de la experiencia y habilidad técnica del explorador. El objetivo del estudio fue revisar y evaluar la calidad metodológica de los trabajos de fiabilidad de la exploración ecográfica sobre el tendón de Aquiles.

**Material y método:** Se realizó una revisión sistemática, siguiendo las recomendaciones PRISMA, buscando en las bases de datos MEDLINE, SPORTDiscus, Academic Search, CINAHL, E-Journals, IBECs, LILACS OVID, DIALNET, COCHRANE y PEDro. Se generó una herramienta formada por 9 ítems para valorar la calidad metodológica de la descripción de las exploraciones ecográficas. Para evaluar la calidad metodológica de los estudios de fiabilidad seleccionados se utilizó la escala QAREL.

**Resultados:** Se localizaron 227 artículos y tras un proceso de selección exhaustivo se incluyeron 13 trabajos. La información proporcionada relativa al cegamiento de los exploradores y al análisis estadístico es insuficiente en un 70% y en un 69% de los estudios analizados respectivamente.

**Conclusiones:** La descripción del método ecográfico es insuficiente para asegurar la reproducibilidad de los estudios revisados. Además, los artículos presentan problemas metodológicos en aspectos relevantes tales como la representatividad de la muestra, el cegamiento de los exploradores y el adecuado análisis estadístico de los datos.

© 2016 MVClinic. Publicado por Elsevier España, S.L.U. Este es un artículo open access distribuido bajo los términos de la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

The Achilles tendon is the thickest and strongest tendon of the body, able to support loads of up to 12.5 times the body weight during gait. Due to the high loads and tensions that it must sustain, it is one of the tendons that most commonly suffers from injuries<sup>1</sup>. At present, tendon alterations and their causes are still a subject of study and discussion.

It is well known that tendinopathies are the most frequent type of injuries in people who practice sports, whether this be on a professional or recreational level<sup>2</sup>. Associated factors have been identified, including age, sex, vascularization deficiencies, genetic aspects and poor flexibility<sup>3</sup>. Clinically, the clearest manifestation is the appearance of pain and, in many cases, functional incapacitation when tissue changes have already occurred in the tendon structure. At the level of ultrasound imaging, the most frequent changes that occur in tendinopathies are tendon thickening, focal or global hypoechogenicity, calcifications close to the insertion on the calcaneus and, on occasion, signs of neovascularization<sup>4</sup>.

From the point of view of the diagnosis and to oversee the progression of the injury, musculoskeletal ultrasound has an excellent cost-benefit relation; it is harmless, non-invasive and enables dynamic patient assessments. Thus, it is not by chance that this is now the technique of choice selected by many health professionals, including physiotherapists<sup>5,6</sup>.

These images are obtained via graphical representation of the ultrasound (US) waves which, after reflecting on the different body tissues, are once again recorded by the transducer and processed by the specialised software of the ultrasound device in order to obtain an ultrasound image<sup>6</sup>.

However, considering the characteristics of the examination, the quality of the image obtained does not solely depend on the appropriate image optimisation, but rather on the anatomic knowledge, experience and skill of the examiner. It is precisely this factor that is most criticised and discussed when ultrasound is used during quantitative studies.

Thus, it is important to consider that, prior to the use of any ultrasound protocol for data collection during research, the reliability of this method must be known and assessed. It is said that an assessment measure (or study methodology) is reliable when it is able to obtain a similar result under identical measurement conditions<sup>7-9</sup>. The reliability of the assessment measure is the basis for it to be considered valid for assessment or diagnosis.

Although there is published research in which the corresponding reliability study has been included, this is not always the case, or it has not been appropriately reported. Therefore, we performed a systematic review on reliability studies for the ultrasound examination of the Achilles tendon in humans. The purpose of the review was to assess the level of description of the parameters that affect the reproducibility of the ultrasound examination and the degree of compliance with the standards for reliability studies.

## Material and methods

### Study design

We performed a systematic review following the PRISMA recommendations<sup>10</sup>. The literature searches were performed between December 2015 and May 2016.

## Documentary sources

The searches were performed in the MEDLINE database via the PUBMED platform, as well as in the following databases: SPORTDiscus, Academic Search, CINAHL, E-Journals via the EBSCO platform, IBECs and LILACS via the BVS platform, aOVID and DIALNET via the virtual platform of the CEU San Pablo University, COCHRANE and PEDro. Furthermore, manual searching was performed in order to locate the greatest number of articles as possible.

## Search strategy

The searches were performed by two independent reviewers (EPO and FDP), and discrepancies were resolved by discussion with another member of the team (JRD), who adapted the search query syntax according to the corresponding database. No filter was used regarding the date since publication, age, language or type of study in order to obtain a more sensitive selection.

### MEDLINE

#1 ("Ultrasonography"[Mesh] OR ultrasonography[TIAB] OR echography[TIAB] OR sonography[TIAB] OR ultrasound\*[TIAB]) AND (reliability[TIAB] OR reproducibility[TIAB] OR "Reproducibility of Results"[Mesh] OR agreement[TIAB] OR Validity [TIAB] OR inter-observer[TIAB] OR interobserver[TIAB] OR intra-observer[TIAB] OR intraobserver[TIAB] OR inter-rater[TIAB] OR interrater[TIAB] OR intra-rater[TIAB] OR intrarater[TIAB] OR intra-examiner[TIAB] OR intraexaminer[TIAB] OR inter-examiner[TIAB] OR interexaminer[TIAB] OR interexplorer[TIAB] OR test-retest[TIAB] OR "Observer Variation"[Mesh]) AND ("Achilles tendon"[TIAB] OR "Achilles Tendon"[Mesh])

### EBSCO

Advanced search for words appearing in the abstract in SPORTDiscus with Full Text, Academic Search Complete, CINAHL Complete, E-Journals:

#2 (ultrasonography OR echography OR sonography OR ultrasound\*) AND (reliability OR reproducibility OR agreement OR Validity OR inter-observer OR interobserver OR intra-observer OR intraobserver OR inter-rater OR interrater OR intra-rater OR intrarater OR intra-examiner OR intraexaminer OR inter-examiner OR interexaminer OR interexplorer OR test-retest) AND (Achilles tendon)

### OVID

Multifield search for words appearing in the abstract:

#3 (ultrasonography OR echography OR sonography OR ultrasound) AND (reliability OR reproducibility OR agreement OR Validity OR inter-observer OR interobserver OR intra-observer OR intraobserver OR inter-rater OR interrater OR intra-rater OR intrarater OR intra-examiner OR intraexaminer OR inter-examiner OR interexaminer OR interexplorer OR test-retest) AND (Achilles tendon)

### Biblioteca Virtual de Salud (BVS)

Word search in IBECs and LILACS:

#4 (ultrasonography OR echography OR sonography OR ultrasound) AND (reliability OR reproducibility OR agreement OR Validity OR inter-observer OR interobserver OR intra-observer OR intraobserver OR inter-rater OR interrater OR

intra-rater OR intrarater OR intra-examiner OR intraexaminer OR inter-examiner OR interexaminer OR interexplorer OR test-retest) AND (Achilles tendon)

### COCHRANE

Advanced search by words in title/abstract/keywords:

#5 (ultrasonography OR echography OR sonography OR ultrasound) AND (reliability OR reproducibility OR agreement OR Validity OR inter-observer OR interobserver OR intra-observer OR intraobserver OR inter-rater OR interrater OR intra-rater OR intrarater OR intra-examiner OR intraexaminer OR inter-examiner OR interexaminer OR interexplorer OR test-retest) AND (Achilles tendon)

### PEDro

Simple search:

#6 ultrasonography AND Achilles tendon  
#7 sonography AND Achilles tendon  
#8 ultrasound\* AND Achilles tendon  
#9 echography AND Achilles tendon

### Dialnet

Search for words in documents:

#10 ultrasonography AND Achilles tendon  
#11 sonography AND Achilles tendon  
#12 ultrasound\* AND Achilles tendon  
#13 echography AND Achilles tendon

## Selection criteria

We only included publications in journals reviewed by both researchers and that studied the reliability of ultrasound of the Achilles Tendon (AT) in living human subjects. We rejected works in which sonoelastography was used which, although based on ultrasound, has other specific characteristics. We also rejected studies that evaluated reliability in qualitative classifications and diagnosis

## Level of compliance of the ultrasound method description

In order to assess the suitability of the descriptions provided regarding the ultrasound method (which directly affects the reproducibility of the studies), an *ad hoc*, non-validated tool was created based on the clinical and research experience of five authors.

Nine items were described, which were coded as follows: 0 points if the item is not described; 1 point if the item is described unclearly or incompletely, and therefore not reproducible; 2 points if the item is clearly described and therefore reproducible (table 1).

## Methodological quality of the reliability studies

In order to evaluate the methodological quality of the reliability studies, the Quality Appraisal for Reliability Studies (QAREL) scale was used<sup>11</sup> which comprises 11 questions (possible responses: yes i.e. complies, no, unclear or non-applicable) grouped into three groups that describe the internal validity (items 3-9), the external validity (items 1, 2 and 10) and appropriateness of the statistical analyses (item 11).

**Table 1** Description of the scale for evaluating reproducibility of the method

Item	Description
<i>Item 1: Subject position</i>	The position of the subject during the examination is described as well as the position of the anatomic region explored.
<i>Item 2: Anatomic region explored</i>	The exact location of the probe upon the structure of the area under study is described.
<i>Item 3: Plane of exploration (section)</i>	The plane or type of section is described and whether or not it was bilateral.
<i>Item 4: Probe characteristics</i>	The type of probe and its frequency is described.
<i>Item 5: Ultrasound gel or similar</i>	The type of gel or similar is described as well as the amount, in the case of ultrasound gel. This section also includes whether the study refers to the amount of pressure applied with the probe over the area explored. This pressure should be the minimum possible in order to avoid deforming the analysed structure.
<i>Item 6: Characteristics of the ultrasound device</i>	Details regarding the model and brand used, as not all ultrasound devices obtain the same image resolution.
<i>Item 7: Image optimization parameters</i>	The description of the adjustments regarding: work frequency, focus and gain curves, which were established to perform the examinations. In particular, the description of these adjustments is very important, as this will directly influence the ultrasound image obtained.
<i>Item 8: Data storage</i>	Evaluation of whether the images obtained in the examinations were saved separately from the ultrasound device for future analysis, and in which format these were saved (e.g.: PNG, JPEG, TIFF). Some formats save images with loss of quality (compression) which affects the quality and the ability to determine specific variables of image analysis.
<i>Item 9: Characteristics of the examiner</i>	Description of who or whom performed the ultrasound examinations and their experience in this field in years.

Recently<sup>12</sup>, the creators of the scale have published values regarding the reliability of the same indicating that nine of the eleven items have an interobserver reliability that is good or moderate, and that two of the items should perhaps be reviewed.

## Results and discussion

### Selection of studies

After locating 227 articles via search strategies and manual searches, 72 papers were rejected by the ZOTERO reference management software as they were duplicated. Of 155 remainder papers, 142 were excluded because they did not conform to the selection criteria (fig. 1). In the case of difficulties or the absence of consensus between the two reviewers, a third reviewer participated in the final decision. As a result, 13 papers were ultimately included in this review<sup>13-25</sup>.

### General characteristics of the studies

All studies included in the review date from 2003 or later, and most of these date from the last 4 years<sup>13-15,20-22,24</sup> which could be explained by the recent awareness of the need for quality research examining reliability, according to the STARD recommendations<sup>26</sup>. This is similar to the PRISMA<sup>27</sup> recommendations for systematic reviews, STROBE<sup>28</sup> for observational studies, or the more consolidated CONSORT statement<sup>29</sup> for clinical trials.

It is important to note that, despite the fact that the AT is one of the most frequently injured tendons, in most of the articles reviewed, the reliability of the ultrasound examination is studied in asymptomatic healthy subjects, except in the article by Sunding et al.<sup>14</sup>, who examined the reliability of ultrasound in asymptomatic subjects. This is an interesting aspect that affects the external validity of the work: reliability studies should be performed on similar samples of the population upon which the analyses are performed<sup>30</sup>.

### Description of the reproducibility of the ultrasound method

Table 2 displays the results obtained for the different items on the ad hoc scale for evaluating the degree of reproducibility of the ultrasound method.

Regarding the position of study subjects, all examinations were performed in the supine position, however the position of the lower limb while performing the AT examination was not clear in 38% of the papers. The position of the ankle and knee should always be described with the maximal detail, because the width or the area of the transverse section of the AT, i.e. two of the ultrasound variables that are most analyzed in ultrasound, depend on the former.

The anatomic region in which the probe was placed to perform both the ultrasound examinations and the selected ultrasound section is clearly described in approximately three quarter of studies (77%). In particular, the anatomic region is very important, as the AT does not have the same characteristics throughout its entire length. For example,

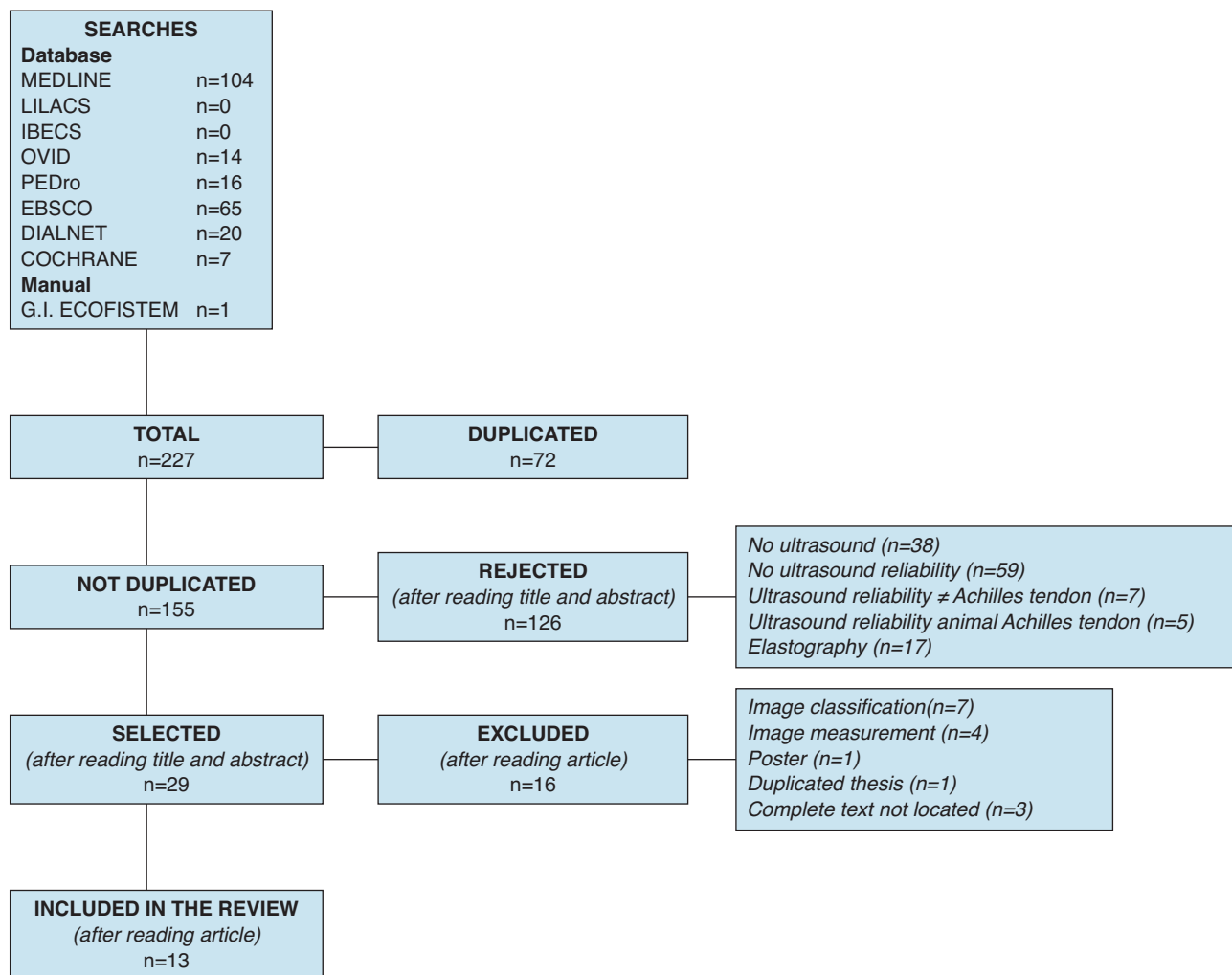


Figure 1 Flowchart of search strategy.

Intziegianni et al.<sup>22</sup> conclude that the measurements taken 2 cm proximal to the calcaneus are less reproducible than those taken at 4 and 6 cm distance.

The type and frequency of the probe is one of the aspects that is best described in all articles (92%); a linear probe

was used in all cases as this is a musculoskeletal examination. Furthermore, in the case of the AT it is important to consider that the area is irregular and concave, and therefore it is impossible to correctly position the probe if it is too long<sup>31</sup>. In this review we have found that the length of the probe is described in 6 studies<sup>13,15,16,20,24,25</sup>.

The amount of gel and the pressure exercised to perform the ultrasound examination is described in only half the studies. It is important to note that it would not make much sense, nor would it be practical, to establish a quantitative amount of gel and, therefore, the ideal amount of gel can be considered as being enough to enable an acoustic coupling to avoid rebound and the dispersion of the ultrasound wave between the transducer and the skin. The authors of the studies should describe, in one way or another, if this was at least considered.

Concerning the pressure applied over the transducer, something similar occurs: the examiner should be aware of the need for exercising the minimal amount of pressure in order to ensure the transduction and in order not to modify the shape of the structure, as this will particularly affect variables such as the thickness or morphological parameters which evaluate the size or roundness of the same.

**Table 2** Frequency of compliance with the items regarding ultrasound methodology explanations

Item	No	Unclear	Yes
1. Subject position	0 (0%)	5 (38.5%)	8 (61.5%)
2. Anatomic region	0 (0%)	3 (23.1%)	10 (76.9%)
3. Ultrasound section	0 (0%)	3 (23.1%)	10 (76.9%)
4. Probe	1 (7.7%)	0 (0%)	12 (92.3%)
5. Gel	7 (53.8%)	5 (38.5%)	1 (7.7%)
6. Device	0 (0%)	0 (0%)	13 (100%)
7. Ultrasound parameters	6 (46.2%)	5 (38.5%)	2 (15.4%)
8. Image	4 (30.8%)	5 (38.5%)	4 (30.8%)
9. Examiner	5 (38.5%)	6 (46.2%)	2 (15.4%)



It is important to note that only Tweedell et al.<sup>21</sup> described the use of gel and the amount of pressure exercised.

The ultrasound device is clearly described in all papers. This item is not the one that most affects the reproducibility of the study, but it allows readers to infer the image obtained and, therefore the results, especially in the cases in which the intensity of the greyscales of the structures are quantified.

The optimization parameters of the ultrasound image are not usually described (46%), or these are performed unclearly (38%), however, these are critical in ultrasound studies that quantify the echogenicity or in textural analyses. Also, it is necessary to specify whether these were changed during the examinations or whether they remained constant, as this will directly influence the image obtained. The parameters were only described clearly in two articles by Sunding et al.<sup>14</sup> and Obst et al.<sup>15</sup>.

Regarding the manipulation of the ultrasound registers, in many studies it is unclear whether the images were saved on external storage devices for future analysis, nor in which format. It is important to consider that there are formats with, and without, quality loss. In the case of performing echogenicity quantifications, lossless image formats are recommended (DICOM, TIFF, BMP)<sup>32</sup>.

The level of experience of the examiner/s is clearly described in approximately one quarter of the studies. Ideally, the years of experience of the examiner in this field should be detailed, which is only the case in two of the articles: Intziagianni et al.<sup>22</sup> and Barfod et al.<sup>24</sup>.

It is worth noting that an overview of the total scores obtained on these types of scales must be performed with precaution as not all items bear the same weight, however, as shown in table 3, none of the evaluated articles obtain the maximal score of 18 points. However, the work by Intziagianni et al.<sup>22</sup> and Obst et al.<sup>15</sup> achieve 15 points.

It seems clear that the methods of ultrasound examination should be detailed. Furthermore, a list of consensual recommendations drafted by experts and specifically for ultrasound reporting would be valuable.

## Methodological quality of reliability studies

Table 4 and figure 2 display the results of the degree of compliance with the QAREL scale.

Regarding the representativeness of the sample, in only 7 studies is this similar to the population upon which these techniques will be performed. Only the article by Sunding et al.<sup>14</sup> evaluates the reliability of ultrasound in patients with symptoms. In our opinion, more reliability studies should be performed in pathological Achilles tendons as it is the most frequent injury seen in daily clinical practice.

In most of the articles, the characteristics of the examiners is not specified, neither regarding their training, nor their years of experience, nor whether they fulfil the same conditions as the examiners responsible for implementing the examination in clinical practice (fig. 2).

The information provided regarding examiner blinding is insufficient, especially when there was more than one examiner. This may compromise the quality of the study with an overstatement of the results based on agreements between examiners which may invalidate the study. The blinding of the examiner regarding the selection criteria and the possible clinical characteristics of the patients should be the maximum possible. In the reviewed works we found that in only 38% of these is the interobserver blindness specified, and in only 23% is intraobserver blindness specified. Concerning the remaining items regarding blinding, virtually 90% do not fulfil these (fig. 2).

The item "control of variable stability" was fulfilled in five of the articles. This item refers to how researchers control the possible changes in the analysed structures during the different measurements performed. Studies must provide explicit details regarding which procedures or guidelines were followed in order to avoid or reduce these possible changes. In four studies that fulfil this criteria, the study subjects were advised not to perform vigorous activities 24-72 hours before the test<sup>15,20,21,23</sup>, and in six articles the subject was told to rest for 10-30 minutes in an attempt to control these changes at the level of the tendon<sup>13,15,20,21,23</sup>.

**Table 3** Data matrix regarding the level of explanation provided for the ultrasound method

Author	Year	I1	I2	I3	I4	I5	I6	I7	I8	I9	Total
Tweedell et al. <sup>29</sup>	2016	2	2	2	2	2	2	1	1	0	14
Intziagianni et al. <sup>25</sup>	2015	2	2	2	2	0	2	1	2	2	15
Barfod et al. <sup>28</sup>	2014	2	2	2	0	1	2	1	0	2	12
Sunding et al. <sup>21</sup>	2014	1	2	1	2	0	2	2	2	1	13
Syha et al. <sup>20</sup>	2014	1	2	1	2	0	2	0	2	1	11
Ryan et al. <sup>24</sup>	2013	2	2	2	2	1	2	1	1	1	14
Obst et al. <sup>22</sup>	2013	2	2	2	2	1	2	2	1	1	15
Kongsgaard et al. <sup>30</sup>	2011	2	2	2	2	1	2	0	1	0	12
Ríos-Díaz et al. <sup>25</sup>	2010	1	2	2	2	0	2	0	2	0	11
Brushøj et al. <sup>17</sup>	2006	1	1	2	2	0	2	0	0	0	8
O'Connor et al. <sup>23</sup>	2004	2	1	1	2	0	2	0	0	1	9
Bjordal et al. <sup>31</sup>	2003	2	2	2	2	0	2	0	1	0	11
Ying et al. <sup>32</sup>	2003	1	1	2	2	1	2	0	0	0	9

I = Items assessed regarding the ultrasound method; 0 = does not comply; 1 = unclear or incomplete; 2 = complies.

**Table 4** Compliance matrix for the questions on the QAREL scale

Author	Year	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
Tweedell et al. <sup>29</sup>	2016	Y	UC	NA	UC	NA	UC	UC	UC	Y	Y	Y
Intziagianni et al. <sup>25</sup>	2015	UC	UC	NA	UC	NA	Y	UC	UC	UC	Y	Y
Basfod et al. <sup>28</sup>	2014	N	Y	Y	Y	NA	UC	UC	UC	UC	Y	UC
Sunding et al. <sup>21</sup>	2014	Y	Y	Y	NA	NA	UC	UC	UC	UC	Y	UC
Syha et al. <sup>20</sup>	2014	N	Y	UC	UC	NA	UC	Y	UC	Y	Y	UC
Ryan et al. <sup>24</sup>	2013	N	UC	NA	UC	NA	UC	UC	UC	Y	Y	UC
Obst et al. <sup>22</sup>	2013	Y	UC	NA	N	NA	UC	UC	N	Y	Y	Y
Kongsgaard et al. <sup>30</sup>	2011	N	UC	NA	UC	NA	UC	UC	UC	UC	Y	N
Ríos-Díaz et al. <sup>25</sup>	2010	Y	Y	Y	Y	NA	UC	Y	UC	UC	Y	Y
Brushøj et al. <sup>17</sup>	2006	Y	UC	UC	Y	Y	UC	UC	UC	UC	N	N
O'Connor et al. <sup>23</sup>	2004	N	Y	Y	UC	NA	UC	UC	S	Y	UC	UC
Bjorðal et al. <sup>31</sup>	2003	Y	Y	NA	UC	NA	UC	UC	UC	UC	UC	N
Ying et al. <sup>32</sup>	2003	Y	UC	Y	NA	NA	UC	UC	UC	UC	Y	UC

P = questions on the QAREL scale; Y = yes, complies; N = no, does not comply; UC = unclear; NA = not applicable.

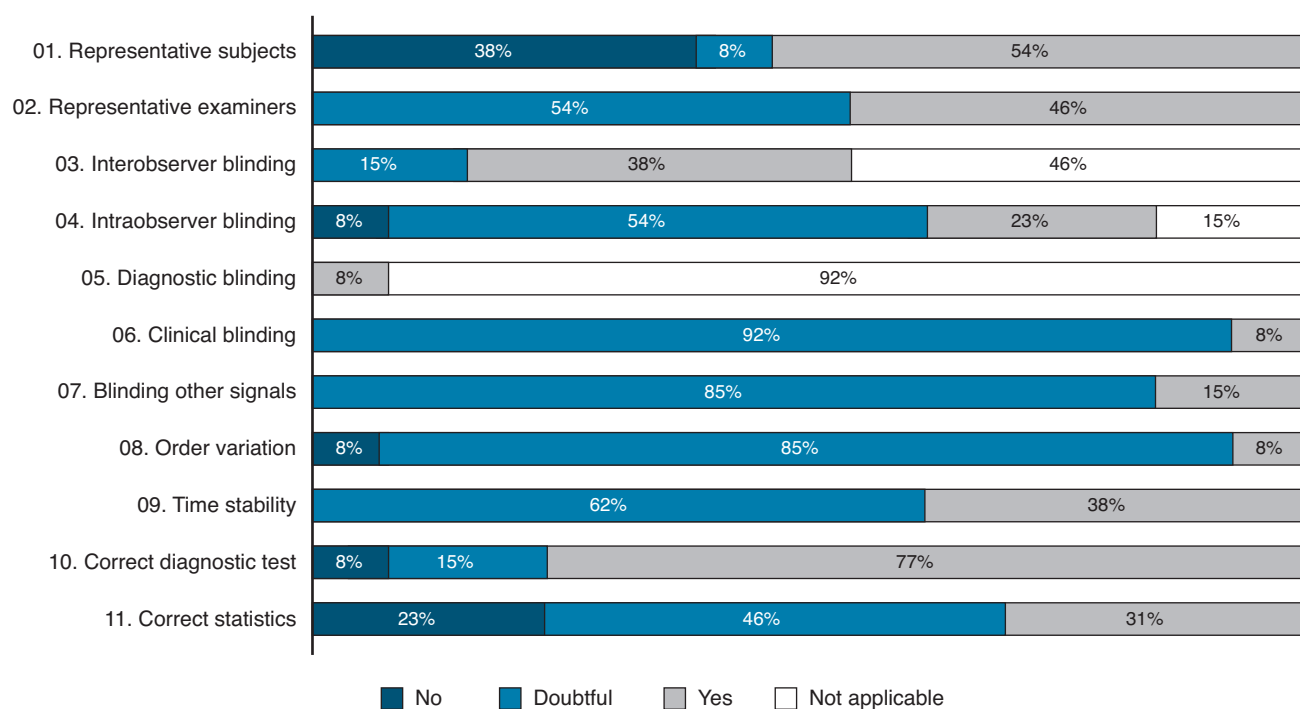
In all studies but one, quantitative ultrasound analysis is performed<sup>13-18,20-25</sup>. In order to determine the reliability of these variables, the authors tend to use the intraclass correlation coefficient (ICC)<sup>25,33</sup>.

Although the use of this coefficient is appropriate, several authors affirm that the study design of reliability studies is frequently incorrect or incomplete<sup>34,35</sup>.

For example, it is important to describe the sample size calculation procedures made for ensuring sufficient statistical power<sup>36,37</sup> and to consider that some of these tests (including the ICC) are based on the assumption that the

normal distribution of data which should be verified and, if this is not the case, the necessary adjustments are required<sup>38,39</sup>.

In the case of the ICC, Shrout and Fleiss (1979)<sup>39</sup> described up to six different ways to summarise this information which should be accurately described in the methods section. Only four of the reviewed papers<sup>15,21,22,25</sup>, explicitly describe the type of model used for the ICC calculations. For example, the authors must indicate if they have calculated the ICC in search of absolute agreement (which would be ideal) or to seek consistency among measurements. Calculations of ICC

**Figure 2** Methodological quality of studies according to the QAREL scale.

are based on the analysis of variance models (ANOVA) and, therefore, it is necessary to indicate if a one or two-way ANOVA has been used and whether the ICC have been calculated based on simple measures or based on the mean of an ensemble of these. Basically, the different models for calculating the ICC may be differentiated regarding how the systematic and random errors are considered in the model equation. This aspect, which may seem banal, has important methodological and theoretical implications. The work by Shout and Fleiss<sup>39</sup>, Weir<sup>38</sup> and McGraw and Wong<sup>40</sup> offer exhaustive descriptions of the implications of performing ICC calculations.

Nonetheless, the ICC is not exempt from criticism<sup>38,40,41</sup> and its uses should be complemented by other indexes, such as the standard error of the mean (SEM) and the minimum detectable difference, both of which are independent from the ICC and have a practical interest as they allow us to know the limits for detecting the changes within the measurement protocol. The visual method described by Bland and Altman (1986)<sup>42</sup> is also considered useful for determining the precision of measurements and to detect possible systematic bias in the records.

In the work by Bjordal et al.<sup>19</sup> and Kongsgaard et al.<sup>16</sup>, the authors use a linear Pearson correlation coefficient as a statistical index of reproducibility. This statistical index should not be used with this aim as, although it indicates consistency between measures, it does not consider the differences that may exist between the repeated measures<sup>33</sup>.

Finally, in order to determine reliability, Brushøj et al.<sup>17</sup> calculated the coefficient of variation (CV). This statistic does not indicate concordance between measures, rather it indicates the variability of the observations compared to the group mean and has many uses, aside from determining the reliability between measurements.

It seems that, despite the belief of some researchers that the ultrasound methods in the examination of the Achilles tendon are perfectly established and valid, this is not the case and most studies analysing the same have serious methodological shortcomings regarding relevant aspects, such as the representativeness of the sample, the blinding of the examiners and the appropriate statistical analysis of the data.

In conclusion, the description of the ultrasound method, in general, is not sufficiently developed, which directly affects the reproducibility of the method and the ability to perform comparisons across studies. It is, thus, recommended to generate consensus guidelines, established by experts, that are specific for ultrasound examinations.

## References

1. Arya S, Kulig K. Tendinopathy alters mechanical and material properties of the Achilles tendon. *J Appl Physiol* 2010;108(3):670-5.
2. Alfredson H. The chronic painful Achilles and Patellar tendon: research on basic biology and treatment. *Scand J Med Sci Sports* 2005;15:252-9.
3. Kader D, Saxena A, Movin T, Maffulli N. Achilles tendinopathy: some aspects of basic science and clinical management. *Br J Sports Med* 2002;36(4):239-49.
4. Jiménez Díaz JF. Diagnóstico clínico y ecográfico de las lesiones del deporte. Murcia: Universidad Católica San Antonio; 2002.
5. Minaya-Muñoz F, Valera-Garrido F, Benito-Domingo A. Ecografía musculoesquelética en fisioterapia. In: Valera F, Minaya F, eds. *Fisioterapia invasiva*. Barcelona: Elsevier; 2013. p. 99-106.
6. Martínez Paya JJ. Anatomía ecográfica del hombro: herramienta de prevención, diagnóstico, investigación y validación de técnicas terapéuticas. Madrid: Editorial Médica Panamericana; 2008.
7. Hurley WL, Denegar CR, Hertel J. Research methods: a framework for evidence-based clinical practice. 1 ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health; 2011.
8. Lexell JE, Downham DY. How to Assess the Reliability of Measurements in Rehabilitation: *Am J Phys Med Rehabil* 2005;84(9):719-23.
9. Charter RA, Feldt LS. Meaning of reliability in terms of correct and incorrect clinical decisions: the art of decision making is still alive. *J Clin Exp Neuropsychol* 2001;23(4):530-7.
10. Urrútia G, Bonfill X. Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis. *Med Clínica* 2010;135(11):507-11.
11. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol* 2010;63(8):854-61.
12. Lucas N, Macaskill P, Irwig L, Moran R, Rickards L, Turner R, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med Res Methodol* 2013; 13(1):1.
13. Syha R, Grau S, Nieß AM, Hein T, Werner MK, Ketelsen D, et al. Computer-based quantification of the Achilles tendon thickness in sequential B-mode ultrasound images: a study of feasibility and reliability. *Arch Orthop Trauma Surg* 2014;134(10):1443-9.
14. Sunding K, Fahlström M, Werner S, Forssblad M, Willberg L. Evaluation of Achilles and patellar tendinopathy with greyscale ultrasound and colour Doppler: using a four-grade scale. *Knee Surg Sports Traumatol Arthrosc* 2016;24(6):1988-96.
15. Obst SJ, Newsham-West R, Barrett RS. In Vivo Measurement of Human Achilles Tendon Morphology Using Freehand 3-D Ultrasound. *Ultrasound Med Biol* 2014;40(1):62-70.
16. Kongsgaard M, Nielsen CH, Hegnsvad S, Aagaard P, Magnusson SP. Mechanical properties of the human Achilles tendon, in vivo. *Clin Biomech* 2011;26(7):772-7.
17. Brushøj C, Henriksen BM, Albrecht-Beste E, Hölmich P, Larsen K, Bachmann Nielsen M. Reproducibility of Ultrasound and Magnetic Resonance Imaging Measurements of Tendon Size. *Acta Radiol* 2006;47(9):954-9.
18. Ying M, Yeung E, Li B, Li W, Lui M, Tsoi C-W. Sonographic evaluation of the size of Achilles tendon: the effect of exercise and dominance of the ankle. *Ultrasound Med Biol* 2003;29(5): 637-42.
19. Bjordal JM, Demmink JH, Ljunggren AE. Tendon thickness and depth from skin for supraspinatus, common wrist and finger extensors, patellar and Achilles tendons: ultrasonography study of healthy subjects. *Physiotherapy* 2003;89(6):375-83.
20. Ryan ED, Rosenberg JG, Scharville MJ, Sobolewski EJ, Thompson BJ, King GE. Test-Retest Reliability and the Minimal Detectable Change for Achilles Tendon Length: A Panoramic Ultrasound Assessment. *Ultrasound Med Biol* 2013;39(12):2488-91.
21. Tweedell AJ, Ryan ED, Scharville MJ, Rosenberg JG, Sobolewski EJ, Kleinberg CR. The influence of ultrasound measurement techniques on the age-related differences in Achilles tendon size. *Exp Gerontol* 2016;76:68-71.
22. Intziogianni K, Cassel M, König N, Müller S, Fröhlich K, Mayer F. Ultrasonography for the assessment of the structural properties of the Achilles tendon in asymptomatic individuals: An intra-rater reproducibility study. *Isokinet Exerc Sci* 2015;23(4): 263-70.
23. O'Connor PJ, Grainger AJ, Morgan SR, Smith KL, Waterton JC, Nash AFP. Ultrasound assessment of tendons in asymptomatic



- volunteers: a study of reproducibility. *Eur Radiol* 2004;14(11): 1968-73.
24. Barfod KW, Riecke AF, Boesen A, Hansen P, Maier JF, Døssing S, et al. Validation of a novel ultrasound measurement of Achilles tendon length. *Knee Surg Sports Traumatol Arthrosc* 2015;23(11):3398-406.
  25. Ríos-Díaz J, Martínez-Payá JJ, del-Baño-Aledo ME, de-Groot-Ferrando A. Fiabilidad y reproducibilidad intra e inter-observador de un método semiautomático de análisis ecográfico del tendón de Aquiles. *Cuest Fisioter* 2010;39(3):190-8.
  26. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Clin Chem* 2003;49:7e18.
  27. Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Med* 2008;6:e1000097.
  28. Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *PLoS Med* 4 2007;e297.
  29. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285:1987-91.
  30. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. In: Knottnerus JA, Buntinx F. *The Evidence Base of Clinical Diagnosis: Theory and Methods of Diagnostic Research*. 2 ed. Oxford: Blackwell Publishing Ltd; 2009. p. 42-62.
  31. Matas RB, Villarrubia MR, Aleu AC. *Ecografía muscular de la extremidad inferior: sistemática de exploración y lesiones en le deporte*. Madrid: Elsevier; 2005.
  32. Ríos-Díaz J, Martínez-Payá JJ, de Groot-Ferrando A, del Baño-Aledo ME. Procesamiento y análisis de imagen en ecografía. In: Valera F, Minaya F, eds. *Fisioterapia Invasiva*. Barcelona: Elsevier; 2013. p. 359-74.
  33. Pita Fernández S, Pérttega Díaz S, Rodríguez Maseda E. La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas. *Cad Aten Primaria* 2003;10(4): 290-6.
  34. English C, Fisher L, Thoires K. Reliability of real-time ultrasound for measuring skeletal muscle size in human limbs in vivo: a systematic review. *Clin Rehabil* 2013;26:934-44.
  35. Zaki R, Bulgiba A, Ismail R, Ismail NA. Statistical Methods Used to Test for Agreement of Medical Instruments Measuring Continuous Variables in Method Comparison Studies: A Systematic Review. *PLoS ONE* 2012;7:e37908.
  36. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101-10.
  37. Shieh G. Optimal sample sizes for the design of reliability studies: Power consideration. *Behav Res Methods* 2014; 46: 772-85.
  38. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res Natl Strength Cond Assoc* 2005;19:231-40.
  39. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420-8.
  40. McGraw KO, Wong S. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1:30-46.
  41. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil* 1998;12:187-99.
  42. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.