# Intra- and Interobserver Reliability of Ultrasound Measurement of the Plantar Fascia

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Journal of Clinical Ultrasound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>JCU-10-091.R1</td>
</tr>
<tr>
<td>Wiley - Manuscript type:</td>
<td>Research Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>foot, musculoskeletal, plantar fascia, Plantar fasciitis, sonography</td>
</tr>
</tbody>
</table>
Ultrasound Measurement of the plantar fascia

Intra- and Interobserver Reliability of Ultrasound Measurement of the Plantar Fascia.
Ultrasound Measurement of the plantar fascia

ABSTRACT

Purpose
To determine intra- and interobserver reliability and measurements precision of Ultrasound Imaging (US) assessment of plantar fascia thickness when using one, the mean of two or three measurements.

Methods
Two experienced observers scanned 20 healthy subjects twice with 60 minutes between test and retest. A GE LOGIQe ultrasound scanner was used in the study. The built-in software in the scanner was used to measure the thickness of the plantar fascia (PF). Reliability was calculated using intraclass correlation coefficient (ICC) and limits of agreement (LOA).

Results
Intraobserver reliability using one sonogram was ICC 0.50 and 0.52, and using the mean of three sonograms intraobserver reliability increased up to 0.77. Interobserver reliability when measuring one sonogram was ICC 0.62 and increased to 0.82 when using the average of three sonograms. LOA showed that when using the average of 3 sonograms, LOA decreased to 0.6 mm, corresponding to 17.5% of the mean thickness of the PF.

Conclusions
The results showed that reliability increases when using the mean of three sonograms compared to one. Limits of agreement based on intratester reliability shows that changes in thickness that are larger than 0.6 mm can be considered actual changes in thickness and not a result of measurement error.

Keywords: foot; musculoskeletal; plantar fascia; plantar fasciitis; reliability; sonography
Ultrasound Measurement of the plantar fascia

Introduction

Ultrasound imaging (US) is a well-established, non-invasive method for examining soft tissue structures of the foot, including the plantar fascia (PF).\textsuperscript{1,2} It is a valuable tool in clinical practice to evaluate the integrity of musculoskeletal structures qualitatively.\textsuperscript{3,4} Healthy tendons are known to have a well-organized, uniform, hyperechoic pattern of collagen.\textsuperscript{5} Conversely, tendons with pathology is often thicker and have a more disorganized, diffuse or hypoechoic appearance on US.\textsuperscript{1,4} In essence, tendon health can be clinically evaluated by recording a grayscale sonogram of the tendon.

Diagnosis of plantar fasciitis is often made due to a combination of patient complaints, localization of pain and measurable thickening of the PF. This has been reported in case control studies where patients with plantar fasciitis was compared to healthy subjects.\textsuperscript{2,7} Interestingly a correlation has been seen between a reduction in the thickness of the PF and a reduction in pain.\textsuperscript{8}

Ultrasonography is often used to support the diagnosis of plantar fasciitis by measuring a thickening of the PF. During rehabilitation US may also be used to secure that rehabilitation is proceeding as expected. The goal during rehabilitation is to observe a reduction in the thickness of the PF as well as a reduction in pain. The main problem with US is that it is a real-time and operator-dependent technique, and factors such as transducer-handling and machine-settings might influence size and appearance of the PF. If patients are examined several times by different examiners, it is essential to know intraobserver and interobserver reliability and precision of the method. This will allow the examiner to know if the changes observed are actual changes in thickness or a result of measurement error. To our knowledge, inter- and intraobserver reliability of US assessment of PF thickness and intra-day variation in the thickness of the PF are not yet known.
Ultrasound Measurement of the plantar fascia

The purpose of the study is to determine intra- and interobserver reliability and measurement precision of US assessment of PF thickness when using one or the average of multiple scans. We expected that intraobserver reliability would be higher than interobserver reliability and that both inter- and intraobserver reliability would increase if the mean of three measurements was used. Overall we expected that through a standardized protocol in which one examiner records and measures three sonograms, US will be reliable (ICC>0.75) and precise (Limits of agreement <20%).

Methods

Participants
Twenty healthy volunteers, aged 20 to 31 years, without lower extremity symptoms, were recruited for this study. Exclusion criteria were prior history of surgery to the foot and current or prior pain in PF. Participants provided consent on forms approved by the local Ethics Committee in Denmark (N-2009-0050).

Examiners
Two experienced observers were recruited for the study. Both had used musculoskeletal US as part of daily examinations for 2-5 years. In order to minimize learning effect from customization to the specific US machine used in the study, both examiners were allowed a recommendation period of 4 sessions with the US machine.

Test
Measurements were carried out using the set of recommendations from the ESSR (European Society of Musculoskeletal Radiology), figure 1. Subjects were positioned in a prone position. The toes were dorsally flexed, and the talo-crural joint was positioned in 0 degrees. As the PF is
Ultrasound Measurement of the plantar fascia

attached at the plantar surface at the toes, dorsal flexion of the toes creates tension in the PF and makes the borders of the fascia more clearly defined. The transducer was placed over the plantar aspect of the hindfoot to examine the PF at the insertion onto the calcaneus. Long-axis sonograms were obtained approximately 0.5 cm medial to the midline of the plantar surface of the foot, figure 2. Sonograms of both the right and the left foot were used in the study. The scanner was from General Electric Healthcare (3135 Easton Tpke, Fairfield, CT, USA), model LOGIQ e. The transducer that was used was a 12L RS, 13 MHz linear transducer. Scan depth was set to 2 cm. Three successive scans were made of each foot to resemble a feasible clinical setting. Between each scan the transducer was moved away from the foot and then repositioned. It was decided at random whether observer 1 or 2 did the first scan and whether the scans were initiated on the left or the right foot. To investigate whether the two examiners would increase reliability during the day of testing, all subjects received a consecutive identification number. This allowed us to ensure that no learning effect would influence the results.

Figure 1 about here.

Figure 2 about here.

Retest

After a period of approximately 60 minutes, the subjects underwent a second ultrasound examination. During the retest, three additional sonograms of the PF were recorded. Care was taken to maintain the same standardized foot position, to keep the ultrasound machine settings constant, and to replicate the exact measurement. As physical activity and time-of-day might influence on the actual thickness of the PF, we chose a period of 60 minutes between test and retest. All subjects where instructed to rest between test and retest.
Ultrasound Measurement of the plantar fascia

Sonogram Analysis

All sonograms were measured using the built-in software in the scanner. PF thickness measurements were made at the insertion onto the calcaneus. A third person randomly loaded sonograms onto the screen for examiner 1 and 2 to measure, thereby blinding the observers to which subject they were measuring.

Data Analysis

Statistical analysis was performed using SPSS version 15.0. The dependant measure of PF thickness was calculated using the built-in software in the scanner. Each sequential measurement was arranged in SPSS as a separate variable, and intraexaminer reliability with 95% confidence intervals (CI) was estimated using two-way mixed-model, consistency-type intraclass correlation coefficients (ICC).

Both examiners were blinded during measurement of the current and previous thickness values. The reliability when using a single measurement was estimated using the first 2 measurement variables and the “single measures” output from SPSS. The reliability when using a mean of 2 measurements was estimated using the first 2 measurement variables and the “average measures” output from SPSS. The reliability when using a mean of 3 measurements was similarly estimated using the first 3 measurement variables and the “average measures” output from SPSS. To determine interobserver reliability, a parallel model (two-way random effect, total agreement) with single measurement and mean values of 2 and 3 measurements was calculated.

Measurement precision was assessed by limits of agreement (LOA) and LOA as a percentage of the mean value (LOA%).
Ultrasound Measurement of the plantar fascia

Results

The sample included in the study consisted of 4 men and 16 women. Mean age was 24.1 (SD 2.6).

Mean body mass index (BMI) was 22.2 (SD 2.1).

Paired t-test shows no significant difference between the right and left PF, (p>0.05), table 1. Paired t-test shows no systematic difference in thickness between test and retest for observer 1 or observer 2 (p>0.05). ICC values show no significant difference in reliability between right and left PF which is the reason why the right and left plantar fascias are pooled in the reliability analysis.

Table 1

Intraobserver

Intraobserver reliability measured by ICC for observer 1 using one sonogram was 0.50 (CI 0.23-0.70) and increased to 0.77 (CI 0.56-0.87) when using the average of three sonograms.

Intraobserver reliability for observer 2 using one sonogram was 0.52 (CI 0.09-0.75) and increased to 0.67 (CI 0.39-0.82) when using the average of three sonograms. When using one sonogram, LOA was 0.8mm for observer 1 and 0.9 for observer 2. LOA decreased to 0.6mm (LOA% of 17.5%) and 0.8mm (LOA% 20.8) when using the mean of three sonograms, table 2.

Table 2

Interobserver

Interobserver reliability when measuring different sonograms measured by ICC was 0.62 (CI 0.37-0.78) and increased to 0.82 (CI 0.65-0.91) when using the mean of three sonograms, table 3. When the observers measured the same set of sonograms, reliability measured by ICC was 0.71 (CI 0.50-0.84) and 0.73 (CI 0.55-0.85) and increased to 0.89 (CI 0.78-0.95) and 0.91 (CI 0.82-0.95)

Table 3
Ultrasound Measurement of the plantar fascia

Doing a visual inspection of the consecutive ID number plotted against the difference between test and retest, there are no sign of learning effect with decreased difference with increasing ID number. Furthermore a Bland-Altman plot with the difference between test and retest plotted against the thickness of the PF shows no sign of increased error with increased thickness of the PF.

Discussion

The purpose of this study was to determine the intra- and interobserver reliability of US assessment of PF thickness and to investigate improvements in measurement precision when using the mean value of three measurements compared to one. Intraobserver reliability using one sonogram was 0.50 and 0.52 and using the mean of three sonograms intraobserver reliability increased to 0.67 and 0.77. Interobserver reliability when using one sonogram was 0.62 and increased to 0.82 when using the average of three sonograms. LOA based on intratester reliability shows that changes in thickness that are larger than 0.8mm can be considered actual changes in thickness and not a result of measurement error.

A study by Liang et al\textsuperscript{12} examined the intraobserver reliability of US to assess the thickness of fascia plantaris in 14 patients diagnosed with plantar fasciitis. They showed ICC values of 0.90 (CI 0.70-0.97). Craig et al\textsuperscript{13} did a similar study on an unknown number of patients with diabetes and found ICC of 0.89. Neither Liang et al\textsuperscript{12} or Craig et al\textsuperscript{13} reported whether the measurements were done by a blinded observer, how long time there was between test and retest or which ICC model was used. This makes a direct comparison difficult.

Intraobserver reliability
Ultrasound Measurement of the plantar fascia

Koppenhaver\textsuperscript{14} examined the intraobserver reliability US to assess the thickness of transverses abdominis (TrA) and multifidi (MU) with a study design similar to this. They found ICC values of 0.83 (CI 0.67-0.91) and 0.79 (CI 0.61-0.81), corresponding to 15.1\% and 19.1\% Standard Error of the Estimate (SEM), which approximately corresponds to LOA of 0.7 mm and 0.9 mm. SEM and LOA are comparable if study designs are similar.\textsuperscript{15} Springer et al\textsuperscript{16} did a similar study where they found a intraobserver SEM of 17\% when measuring the thickness of TrA, corresponding to measurements precision of 1.3 mm. Bjorland\textsuperscript{17} examined intraobserver reliability when measuring the thickness of the tendon from m supraspinatus. They found LOA of 0.4 mm and a mean value of tendon thickness of 6.6 mm.

\textbf{Interobserver reliability}

O’Connor\textsuperscript{18} did an interobserver study of US measurement of Achilles tendon thickness. They found a LOA\% of 29\%. Ying\textsuperscript{19} did a similar study and found interobserver measurement error between 5 observers of 0.9-1.1 mm, which corresponds to a LOA\% of approximately 20\%.

It appears that intraobserver reliability of US assessments of human tendon and muscle thickness are between 0.4 mm and 1.3 mm when measuring healthy tendons. LOA\% is between 15 and 19\%. Interobserver reliability is slightly lower with LOA\% of approximately 20\%. We found intraobserver reliability of ICC 0.77 and 0.67 and LOA 0.6 mm and 0.8 mm, corresponding to 17.5\% and 20.8\%. Interobserver reliability was ICC 0.82, LOA 0.57, corresponding to 15.8\%. Both intra- and interobserver reliability found in the literature correspond to the findings in this study.

\textbf{Clinical Relevance}
Ultrasound Measurement of the plantar fascia

Genc et al\textsuperscript{26} showed a significant correlation between a decrease in thickness of the PF and reduction in pain. This was present at both 1 and 6 months after treatment. They found that PF thickness decreased 1.7 mm from initial treatment until follow-up at 1 month. At the next follow-up after 6 months, the PF thickness had decreased another 0.6 mm. Put into context of the current study, the change in PF thickness that they observed in the first month of treatment reflects actual changes in the thickness. However the change in thickness between 1 and 6 months of follow-up might be caused by measurement error as they only recorded one sonogram. The results of this study showed that if only one sonogram is measured, then the minimal detectable change in PF thickness is 0.8 mm while it decreases to 0.6 mm if three sonograms are used.

Limitations of ICC

In order to analyze intra- and interobserver reliability, we used ICC and LOA. ICC is a very common method for describing reliability in different settings. However ICC has some limitations. The size of ICC depends on the ratio of the variability between subjects to the total variability.\textsuperscript{18} This means that the size of ICC may be affected by factors related to the study sample itself.\textsuperscript{18} If two separate test-retest studies are performed on two study samples where one study sample consists of a homogenous group and the other consists of a very heterogeneous groups, the results will be different. This is not because the methods being used have different reliability in the two groups, but because ICC is dependent on the ratio of the variability between subjects to the total variability.\textsuperscript{18} The higher ICC in a heterogeneous group is due to the calculations that ICC is based on and not due to the fact that the methods in the two study samples disagree. The consequence is that it is easier to disguise measurement error in a heterogeneous population. Liang et al\textsuperscript{12} and Craig et al\textsuperscript{13} both examined pathological groups that may consist of a more heterogeneous sample. This
Ultrasound Measurement of the plantar fascia may explain why the LOA found in the current study corresponds to other studies, but ICC in the current study is lower than what was reported by Liang et al\textsuperscript{12} and Craig et al.\textsuperscript{13}

**Comparison of mean thickness of the PF**

The earliest study where the thickness of the Pf was quantified by US was published in 1993.\textsuperscript{21} Studies that examine healthy subjects find mean values of the thickness between 2.6 and 3.9 mm.\textsuperscript{7, 21-27} The two largest studies, which included 154 and 110 subjects, found mean thickness of the PF of 3.2 mm (women and men combined) and 3.3 mm (±0.3 mm) for women and 3.9 mm (±0.5 mm) for men. The average values in these studies are very similar to the ones found in the present study. Discrepancy in the positioning of subjects, different reference point of measurement and body mass index may to some extent explain the variation from 2.6 to 3.9 mm.

**Largest measurement error from scan or measurement**

The results in table 2 and 3 show that the interobserver reliability when both observers measure the same scans is higher than when they measure separate scans. This indicates that scan is a source of error rather than measurement of the scan. The thickness of the PF is not identical across the entire insertion onto the calcaneus.\textsuperscript{28} This means that movement of the transducer might affect the recorded thickness. Furthermore it indicates that some of the variance across the different scan is not only due to measurement error, but reflects actual changes in thickness.

**Conclusion**

Measurement of PF thickness by US is a reliable method. If the same observer measures the same person twice, the difference will be maximum 0.6 - 0.8 mm (ICC 0.67 – 0.77) in 95 % of the examinations when using the mean of three sonograms. If two different observers examine the same
subject twice, the measurement error will be maximum 0.6 mm (ICC 0.82) in 95% of the examinations.
Ultrasound Measurement of the plantar fascia

Reference List


Ultrasound Measurement of the plantar fascia


Ultrasound Measurement of the plantar fascia


Ultrasound Measurement of the plantar fascia

23. Pascual HJ, Garcia JM, Matamoros EC, et al. Relationship of body mass index, ankle
dorsiflexion, and foot pronation on plantar fascia thickness in healthy, asymptomatic

Imaging 2003; 27: 353-357

25. Ozdemir H, Yilmaz E, Murat A, et al. Sonographic evaluation of plantar fasciitis and


aponeurosis: cadaveric study using ultrasonography and magnetic resonance
imaging. Skeletal Radiol 2008; 37: 929-935
Ultrasound Measurement of the plantar fascia

Intra- and Interobserver Reliability of Ultrasound Measurement of the Plantar Fascia.
ABSTRACT

Purpose

To determine intra- and interobserver reliability and measurements precision of Ultrasound Imaging (US) assessment of plantar fascia thickness when using one, the mean of two or three measurements.

Methods

Two experienced observers scanned 20 healthy subjects twice with 60 minutes between test and retest. A GE LOGIQe ultrasound scanner was used in the study. The built-in software in the scanner was used to measure the thickness of the plantar fascia (PF). Reliability was calculated using intraclass correlation coefficient (ICC) and limits of agreement (LOA).

Results

Intraobserver reliability using one sonogram was ICC 0.50 and 0.52, and using the mean of three sonograms intraobserver reliability increased up to 0.77. Interobserver reliability when measuring one sonogram was ICC 0.62 and increased to 0.82 when using the average of three sonograms.

LOA showed that when using the average of 3 sonograms, LOA decreased to 0.6 mm, corresponding to 17.5% of the mean thickness of the PF.

Conclusions

The results showed that reliability increases when using the mean of three sonograms compared to one. Limits of agreement based on intratester reliability shows that changes in thickness that are larger than 0.6 mm can be considered actual changes in thickness and not a result of measurement error.

Keywords: foot; musculoskeletal; plantar fascia; plantar fasciitis; reliability; sonography
Ultrasound Measurement of the plantar fascia

Introduction

Ultrasound imaging (US) is a well-established, non-invasive method for examining soft tissue structures of the foot, including the plantar fascia (PF). It is a valuable tool in clinical practice to evaluate the integrity of musculoskeletal structures qualitatively. Healthy tendons are known to have a well-organized, uniform, hyperechoic pattern of collagen. Conversely, tendons with pathology are often thicker and have a more disorganized, diffuse or hypoechoic appearance on US. In essence, tendon health can be clinically evaluated by recording a grayscale sonogram of the tendon.

Diagnosis of plantar fasciitis is often made due to a combination of patient complaints, localization of pain and measurable thickening of the PF. This has been reported in case control studies where patients with plantar fasciitis were compared to healthy subjects. Interestingly a correlation has been seen between a reduction in the thickness of the PF and a reduction in pain.

Ultrasonography is often used to support the diagnosis of plantar fasciitis by measuring a thickening of the PF. During rehabilitation US may also be used to secure that rehabilitation is proceeding as expected. The goal during rehabilitation is to observe a reduction in the thickness of the PF as well as a reduction in pain. The main problem with US is that it is a real-time and operator-dependent technique, and factors such as transducer-handling and machine-settings might influence size and appearance of the PF. If patients are examined several times by different examiners, it is essential to know intraobserver and interobserver reliability and precision of the method. This will allow the examiner to know if the changes observed are actual changes in thickness or a result of measurement error. To our knowledge, inter- and intraobserver reliability of US assessment of PF thickness and intra-day variation in the thickness of the PF are not yet known.
The purpose of the study is to determine intra- and interobserver reliability and measurement precision of \textit{US} assessment of \textit{PF} thickness when using one or the average of multiple scans. We expected that intraobserver reliability would be higher than interobserver reliability and that both inter- and intraobserver reliability \textit{would} increase if the mean of three measurements was used. Overall we expected that through a standardized protocol in which one examiner records and measures three \textit{sonograms}, \textit{US} will be reliable (ICC$>0.75$) and precise (Limits of agreement $<20\%$).

**Methods**

**Participants**

Twenty healthy volunteers, aged 20 to 31 years, without lower extremity symptoms, were recruited for this study. Exclusion criteria were prior history of surgery to the foot \textit{and} current or prior pain in \textit{PF}. Participants provided consent on forms approved by the \textit{local} Ethics Committee in Denmark (N-2009-0050).

**Examiners**

Two experienced observers were recruited for the study. Both \textit{had} used \textit{musculoskeletal US} as part of daily examinations for 2-5 years. \textit{In order to minimize learning effect from customization to the specific US machine used in the study}, both examiners were allowed a recommendation period of 4 sessions with the \textit{US} machine.

**Test**

Measurements were carried out using the set of recommendations from the ESSR (European Society of Musculoskeletal Radiology), \textit{figure 1}. Subjects were positioned in a prone position. The toes were dorsally flexed, and the talo-crural joint was positioned in 0 degrees. As the \textit{PF} is...
Ultrasound Measurement of the plantar fascia

Attached at the plantar surface at the toes, dorsal flexion of the toes creates tension in the PF and makes the borders of the fascia more clearly defined. The transducer was placed over the plantar aspect of the hindfoot to examine the PF at the insertion onto the calcaneus. Long-axis sonograms were obtained approximately 0.5 cm medial to the midline of the plantar surface of the foot, figure 2. Sonograms of both the right and the left foot were used in the study. The scanner was from General Electric Healthcare (3135 Easton Tpke, Fairfield, CT, USA), model LOGIQ e. The transducer that was used was a 12L RS, 13 MHz linear transducer. Scan depth was set to 2 cm.

Three successive scans were made of each foot to resemble a feasible clinical setting. Between each scan the transducer was moved away from the foot and then repositioned. It was decided at random whether observer 1 or 2 did the first scan and whether the scans were initiated on the left or the right foot. To investigate whether the two examiners would increase reliability during the day of testing, all subjects received a consecutive identification number. This allowed us to ensure that no learning effect would influence the results.

Figure 1 about here.

Figure 2 about here.

Retest

After a period of approximately 60 minutes, the subjects underwent a second ultrasound examination. During the retest, three additional sonograms of the PF were recorded. Care was taken to maintain the same standardized foot position, to keep the ultrasound machine settings constant, and to replicate the exact measurement. As physical activity and time-of-day might influence on the actual thickness of the PF, we chose a period of 60 minutes between test and retest. All subjects were instructed to rest between test and retest. No significant change in the quantitative ultrasound outcomes were expected during the initial and final tests, because subjects were instructed to rest between these tests.
Ultrasound Measurement of the plantar fascia

Sonogram Analysis

All sonograms were measured using the built-in software in the scanner. PF thickness measurements were made at the insertion onto the calcaneus. A third person randomly loaded sonograms onto the screen for examiner 1 and 2 to measure, thereby blinding the observers to which subject they were measuring.

Data Analysis

Statistical analysis was performed using SPSS version 15.0. The dependant measure of PF thickness was calculated using the built-in software in the scanner. Each sequential measurement was arranged in SPSS as a separate variable, and intraexaminer reliability with 95% confidence intervals (CI) was estimated using two-way mixed-model, consistency-type intraclass correlation coefficients (ICC).

Both examiners were blinded during measurement of the current and previous thickness values. The reliability when using a single measurement was estimated using the first 2 measurement variables and the “single measures” output from SPSS. The reliability when using a mean of 2 measurements was estimated using the first 2 measurement variables and the “average measures” output from SPSS. The reliability when using a mean of 3 measurements was similarly estimated using the first 3 measurement variables and the “average measures” output from SPSS. To determine interobserver reliability, a parallel model (two-way random effect, total agreement) with single measurement and mean values of 2 and 3 measurements was calculated.

Measurement precision was assessed by limits of agreement (LOA) and LOA as a percentage of the mean value (LOA%).
Ultrasound Measurement of the plantar fascia

Results

The sample included in the study consisted of 4 men and 16 women. Mean age was 24.1 (SD 2.6).

Mean body mass index (BMI) was 22.2 (SD 2.1).

Paired t-test shows no significant difference between the right and left PF (p>0.05), table 1. Paired t-test shows no systematic difference in thickness between test and retest for observer 1 or observer 2 (p>0.05). ICC values show no significant difference in reliability between right and left PF which is the reason why the right and left plantar fascias are pooled in the reliability analysis.

Table 1

Intraobserver

Intraobserver reliability measured by ICC for observer 1 using one sonogram was 0.50 (CI 0.23-0.70) and increased to 0.77 (CI 0.56-0.87) when using the average of three sonograms. Intraobserver reliability for observer 2 using one sonogram was 0.52 (CI 0.09-0.75) and increased to 0.67 (CI 0.39-0.82) when using the average of three sonograms. When using one sonogram, LOA was 0.8mm for observer 1 and 0.9 for observer 2. LOA decreased to 0.6mm (LOA% of 17.5%) and 0.8mm (LOA% 20.8) when using the mean of three sonograms, table 2.

Table 2

Interobserver

Interobserver reliability when measuring different sonograms measured by ICC was 0.62 (CI 0.37-0.78) and increased to 0.82 (CI 0.65-0.91) when using the mean of three sonograms, table 3. When the observers measured the same set of sonograms, reliability measured by ICC was 0.71 (CI 0.50-0.84) and 0.73 (CI 0.55-0.85) and increased to 0.89 (CI 0.78-0.95) and 0.91 (CI 0.82-0.95)
Ultrasound Measurement of the plantar fascia

Doing a visual inspection of the consecutive ID number plotted against the difference between test and retest, there are no sign of learning effect with decreased difference with increasing ID number. Furthermore a Bland-Altman plot with the difference between test and retest plotted against the thickness of the PF shows no sign of increased error with increased thickness of the PF.

Discussion

The purpose of this study was to determine the intra- and interobserver reliability of US assessment of PF thickness and to investigate improvements in measurement precision when using the mean value of three measurements compared to one. Intraobserver reliability using one sonogram was 0.50 and 0.52 and using the mean of three sonograms intraobserver reliability increased to 0.67 and 0.77. Interobserver reliability when using one sonogram was 0.62 and increased to 0.82 when using the average of three sonograms. LOA based on intratester reliability shows that changes in thickness that are larger than 0.8 mm can be considered actual changes in thickness and not a result of measurement error.

A study by Liang et al12 examined the intraobserver reliability of US to assess the thickness of fascia plantaris in 14 patients diagnosed with plantar fasciitis. They showed ICC values of 0.90 (CI 0.70-0.97). Craig et al13 did a similar study on an unknown number of patients with diabetes and found ICC of 0.89. Neither Liang et al12 or Craig et al13 reported whether the measurements were done by a blinded observer, how long time there was between test and retest or which ICC model was used. This makes a direct comparison difficult.

Intraobserver reliability
Koppenhaver examined the intraobserver reliability US to assess the thickness of transverses abdominis (TrA) and multifidi (MU) with a study design similar to this. They found ICC values of 0.83 (CI 0.67-0.91) and 0.79 (CI 0.61-0.81), corresponding to 15.1% and 19.1% Standard Error of the Estimate (SEM), which approximately corresponds to LOA of 0.7 mm and 0.9 mm. SEM and LOA are comparable if study designs are similar. Springer et al did a similar study where they found a intraobserver SEM of 17% when measuring the thickness of TrA, corresponding to measurements precision of 1.3 mm. Bjordal examined intraobserver reliability when measuring the thickness of the tendon from m supraspinatus. They found LOA of 0.4 mm and a mean value of tendon thickness of 6.6 mm.

Interobserver reliability

O’Connor did an interobserver study of US measurement of Achilles tendon thickness. They found a LOA% of 29%. Ying did a similar study and found interobserver measurement error between 5 observers of 0.9-1.1 mm, which corresponds to a LOA% of approximately 20%.

It appears that intraobserver reliability of US assessments of human tendon and muscle thickness are between 0.4 mm and 1.3 mm when measuring healthy tendons. LOA% is between 15 and 19%. Interobserver reliability is slightly lower with LOA% of approximately 20%. We found intraobserver reliability of ICC 0.77 and 0.67 and LOA 0.6 mm and 0.8 mm, corresponding to 17.5% and 20.8%. Interobserver reliability was ICC 0.82, LOA 0.57, corresponding to 15.8%. Both intra- and interobserver reliability found in the literature correspond to the findings in this study.

Clinical Relevance
Ultrasound Measurement of the plantar fascia

Gene et al. showed a significant correlation between a decrease in thickness of the PF and reduction in pain. This was present at both 1 and 6 months after treatment. They found that PF thickness decreased 1.7 mm from initial treatment until follow-up at 1 month. At the next follow-up after 6 months, the PF thickness had decreased another 0.6 mm. Put into context of the current study, the change in PF thickness that they observed in the first month of treatment reflects actual changes in the thickness. However, the change in thickness between 1 and 6 months of follow-up might be caused by measurement error as they only recorded one sonogram. The results of this study showed that if only one sonogram is measured, then the minimal detectable change in PF thickness is 0.8 mm while it decreases to 0.6 mm if three sonograms are used.

Limitations of ICC

In order to analyze intra- and interobserver reliability, we used ICC and LOA. ICC is a very common method for describing reliability in different settings. However, ICC has some limitations. The size of ICC depends on the ratio of the variability between subjects to the total variability. If two separate test-retest studies are performed on two study samples where one study sample consists of a homogenous group and the other consists of a very heterogeneous group, the results will be different. This is not because the methods being used have different reliability in the two groups, but because ICC is dependent on the ratio of the variability between subjects to the total variability. The higher ICC in a heterogeneous group is due to the calculations that ICC is based on and not due to the fact that the methods in the two study samples disagree. The consequence is that it is easier to disguise measurement error in a heterogeneous population. Liang et al. and Craig et al. both examined pathological groups that may consist of a more heterogeneous sample. This
Ultrasound Measurement of the plantar fascia

may explain why the LOA found in the current study corresponds to other studies, but ICC in the
current study is lower than what was reported by Liang et al.\textsuperscript{12} and Craig et al.\textsuperscript{13}

**Comparison of mean thickness of the PF**

The earliest study where the thickness of the PF was quantified by US was published in 1993.\textsuperscript{21}

Studies that examine healthy subjects find mean values of the thickness between 2.6 and 3.9 mm.\textsuperscript{7}

\textsuperscript{21-27} The two largest studies, which included 154 and 110 subjects, found mean thickness of the PF
of 3.2 mm (women and men combined) and 3.3 mm(±0.3 mm) for women and 3.9 mm (±0.5 mm)

for men. The average values in these studies are very similar to the ones found in the present study.

Discrepancy in the positioning of subjects, different reference point of measurement and body mass
index may to some extent explain the variation from 2.6 to 3.9 mm.

**Largest measurement error from scan or measurement**

The results in table 2 and 3 show that the interobserver reliability when both observers measure the
same scans is higher than when they measure separate scans. This indicates that scan is a source of
error rather than measurement of the scan. The thickness of the PF is not identical across the entire
insertion onto the calcaneus.\textsuperscript{28} This means that movement of the transducer might affect the
recorded thickness. Furthermore it indicates that some of the variance across the different scan is
not only due to measurement error, but reflects actual changes in thickness.

**Conclusion**

Measurement of PF thickness by US is a reliable method. If the same observer measures the same
person twice, the difference will be maximum 0.6 - 0.8 mm (ICC 0.67 – 0.77) in 95 % of the
examinations when using the mean of three sonograms. If two different observers examine the same
Ultrasound Measurement of the plantar fascia

subject twice, the measurement error will be maximum 0.6 mm (ICC 0.82) in 95% of the examinations.

Clinical Relevance

The results showed that reliability increases when using the mean of three images compared to one. Limits of agreement based on intrater reliability shows that changes in diameter that are larger than 0.63 mm can be considered actual changes in diameter and not a result of measurement error.
Reference List


Ultrasound Measurement of the plantar fascia

muscles, gender, body mass index, and hand dominance. J Orthop Sports Phys Ther
2006; 36: 289-297

17. Bjordal, JM, Demmink, JH, and Ljunggren AE. Tendon Thickness and Depth from Skin for
Supraspinatus, Common Wrist and Finger Extensors, Patellar and Achilles Tendons:

asymptomatic volunteers: a study of reproducibility. Eur Radiol 2004; 14: 1968-
1973

effect of exercise and dominance of the ankle. Ultrasound Med Biol 2003; 29: 637-
642

20. Norman G The pruning of Doctor McDonough and the "humanizing" of statistics. Fertil
Steril 1997; 68: 570-573

21. Wall JR, Harkness MA, Crawford A Ultrasound diagnosis of plantar fasciitis. Foot Ankle
1993; 14: 465-470

bands. A retrospective study of 211 symptomatic feet. J Am Podiatr Med Assoc
2002; 92: 444-449
23. Pascual HJ, Garcia JM, Matamoros EC, et al. Relationship of body mass index, ankle
dorsiflexion, and foot pronation on plantar fascia thickness in healthy, asymptomatic

Imaging 2003; 27: 353-357

25. Ozdemir H, Yilmaz E, Murat A, et al. Sonographic evaluation of plantar fasciitis and


aponeurosis: cadaveric study using ultrasonography and magnetic resonance
imaging. Skeletal Radiol 2008; 37: 929-935
<table>
<thead>
<tr>
<th>Diameter of fascia plantaris [mm]</th>
<th>Men</th>
<th>Women</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>Left</td>
<td>Right</td>
<td>Left</td>
</tr>
<tr>
<td>Diameter observer 1</td>
<td>3.93 (0.14)</td>
<td>3.82 (0.16)</td>
<td>3.44 (0.36)</td>
</tr>
<tr>
<td>Diameter observer 2</td>
<td>4.03 (0.39)</td>
<td>3.93 (0.39)</td>
<td>3.54 (0.37)</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th>Limits of agreement (LOA)</th>
<th>1 image</th>
<th>% of observers average</th>
<th>Mean of 2 images.</th>
<th>% of observers average</th>
<th>Mean of 3 images.</th>
<th>% of observers average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 1 [mm]</td>
<td>0.76</td>
<td>21.1 %</td>
<td>0.67</td>
<td>18.6 %</td>
<td>0.63</td>
<td>17.5 %</td>
</tr>
<tr>
<td>Observer 2 [mm]</td>
<td>0.91</td>
<td>25.3 %</td>
<td>0.82</td>
<td>22.8 %</td>
<td>0.75</td>
<td>20.8 %</td>
</tr>
<tr>
<td>Interobserver [mm]</td>
<td>0.67</td>
<td>18.6 %</td>
<td>0.62</td>
<td>17.2 %</td>
<td>0.57</td>
<td>15.8 %</td>
</tr>
<tr>
<td>Interobserver, both observers measure observer 1’s scans. [mm]</td>
<td>0.59</td>
<td>16.3 %</td>
<td>0.52</td>
<td>14.4 %</td>
<td>0.47</td>
<td>13.0 %</td>
</tr>
<tr>
<td>Interobserver, both observers measure observer 2’s scans. [mm]</td>
<td>0.55</td>
<td>15.6 %</td>
<td>0.49</td>
<td>13.9 %</td>
<td>0.43</td>
<td>12.2 %</td>
</tr>
</tbody>
</table>
Table 3

<table>
<thead>
<tr>
<th>Intraclass correlation coefficient.</th>
<th>1 image</th>
<th>Mean of 2 images</th>
<th>Mean of 3 images.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intraobserver, observer 1 [ICC 95 % CI]</td>
<td>0.50 (0.23-0.70),</td>
<td>0.73 (0.48-0.85)</td>
<td>0.77 (0.56-0.87)</td>
</tr>
<tr>
<td>Intraobserver, observer 2 [ICC 95 % CI]</td>
<td>0.52 (0.09-0.75)</td>
<td>0.64 (0.32-0.81)</td>
<td>0.67 (0.39-0.82)</td>
</tr>
<tr>
<td>Interobserver [ICC 95 % CI]</td>
<td>0.62 (0.37-0.78)</td>
<td>0.80 (0.60-0.90)</td>
<td>0.82 (0.65-0.91)</td>
</tr>
<tr>
<td>Interobserver [ICC 95 % CI]. Both observers measure observer one`s scans.</td>
<td>0.71 (0.50-0.84)</td>
<td>0.86 (0.71-0.93)</td>
<td>0.89 (0.78-0.95)</td>
</tr>
<tr>
<td>Interobserver [ICC 95 % CI]. Both observers measure observer two`s scans.</td>
<td>0.73 (0.55-0.85)</td>
<td>0.89 (0.79-0.94)</td>
<td>0.91 (0.82-0.95)</td>
</tr>
</tbody>
</table>
Table legends

Table 1: Diameter of the plantar fascia sub grouped by gender as well as combined, shown for observer 1 and 2.

Table 2: Intra- and interobserver reliability presented as Limits of Agreement (LOA), and LOA as percentage of observers’ average diameter based on all measurements.

Table 3: Intra- and interobserver reliability. Values are presented as ICC with 95% confidence interval.
Figure 1 Long-axis sonograms of the plantar fascia at the insertion onto the calcaneus.
254x172mm (150 x 150 DPI)
Figure 2 The blue bar on the image indicates placement of the probe on the plantar aspect of the hindfoot.

96x155mm (150 x 150 DPI)