Reliability and Interpretability of Sonographic Measurements of Palmar Dupuytren Nodules

Molenkamp, Sanne; van Straalen, Roel J M; Werker, Paul M. N.; Broekstra, Dieuwke

Published in:
The Journal of Hand Surgery

DOI:
10.1016/j.jhsa.2020.01.005

Publication date:
2020

EDITOR'S CHOICE

Reliability and Interpretability of Sonographic Measurements of Palmar Dupuytren Nodules

Sanne Molenkamp, MD,* Roel J. M. van Straalen, MD,* Paul M. N. Werker, MD, PhD,* Dieuwke C. Broekstra, MSc, PhD*

Purpose In the future, it is expected that treatment of Dupuytren disease (DD) may shift toward control of early disease. Ultrasound might be an accurate method to measure the outcome of such treatment. The aim of this study was to assess the reliability of sonographic measurement of palmar nodules.

Methods Fifty patients with nodules characteristic for early disease were assessed with ultrasound by 2 observers. Four different aspects of DD nodules were measured in the transversal and sagittal planes, width, depth, circumference, and area. The intra- and interobserver reliabilities were calculated using the intraclass correlation coefficient (ICC). The standard error of measurement (SEM) and the smallest detectable change (SDC) were also calculated for each aspect.

Results The intraobserver reliability was good (ICC, 0.724 [0.562−0.833] to 0.886 [0.808−0.934]), except for width in the sagittal direction (ICC, 0.671 [0.484−0.799]). The interobserver reliability was moderate (ICC, 0.385 [0.126−0.596] to 0.757 [0.538−0.869]). The intraobserver ICCs of area were highest (transverse, 0.847 [0.744−0.893]; sagittal, 0.886 [0.808−0.934]). The SEM and SDC of area were 6.1 and 16.9 mm^2 in the transverse and 8.0 and 22.2 mm^2 in the sagittal plane.

Conclusions The intraobserver reliability of sonographic assessment of DD nodules is good. The measurement of area is the most reliable and is, therefore, recommended for future studies. However, even single-observer measurements have a clear dispersion, and a change beyond 16.9 (61%) and 22.2 mm^2 (79%) has to be observed in the transverse and sagittal planes, respectively, before it can be considered as regression or progression.

Clinical relevance Repeated ultrasonographic measurements in DD should ideally be done by a single observer, using area of the nodule in the sagittal plane. Change beyond 16.9 (transverse) and 22.2 (sagittal) mm2 can be considered as a real change in nodule size. *(J Hand Surg Am. 2020;45(6):488−494.*

Key words Dupuytren contracture, nodule size, observer variation, reproducibility of results, ultrasonography.

Additional Material at jhandsurg.org

*From the *Department of Plastic Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.*

Received for publication January 27, 2019; accepted in revised form January 10, 2020.

No benefits in any form have been received or will be received related directly or indirectly to the subject of this article.

Corresponding author: Sanne Molenkamp, MD, Department of Plastic Surgery, University Medical Center Groningen, BB81, Hanzeplein 1, 9713 GZ Groningen, The Netherlands; e-mail: s.molenkamp@umcg.nl.

0363-5023/20/4506-0003$36.00/0
https://doi.org/10.1016/j.jhsa.2020.01.005

In view of recent findings in Dupuytren disease (DD) research, it is possible that, in the future, patients may undergo medical treatments aiming at disease control or even regression, instead of current treatment modalities aimed at symptom relief.[1−3] The outcome of these therapies is more difficult to measure because these patients may not have the characteristic contractures yet. Several outcome measures have been proposed in the literature, of which measurement of area of disease using a tumorimeter has

been investigated most thoroughly.[4] However, with this method, the projection of a nodule is measured on the overlying skin and measurements can only be performed in the frontal plane. Some studies have described the use of ultrasound (US) for the measurement and follow-up of DD nodules.[5,6] With US, the actual size of a nodule can be measured, instead of a projection on the overlying skin, and measurements can be performed in the sagittal and transverse planes. However, US is a dynamic tool and measurements may be influenced by parameters like probe direction and amount of pressure applied by the investigator.[7] Furthermore, it is uncertain whether the borders of DD tissue are always clear, owing to DD fibers extending to the skin and underlying structures. Finally, different types of echogenicity have been observed in DD tissue, which can also complicate the identification of the borders of a nodule.[6] No previous research has studied the reliability of US for the measurement of disease extent in patients with early DD. The aim of this study was to assess the reliability of sonographic measurement of palmar DD nodules by calculating the intra- and interobserver agreement. We also aimed to define the minimum changes in nodule size that have to be observed before they can be interpreted as progression or regression, by calculating the standard error of measurement (SEM) and the smallest detectable change (SDC).

METHODS

Patients

Patients with palmar DD nodules were asked to participate. The study was approved by the local Medical Ethics Committee and all participants gave written informed consent. All participants visited the outpatient clinic of the Department of Plastic Surgery between June 2016 and July 2017 and underwent sonographic evaluation of 1 palpable palmar DD nodule that was not clinically part of a cord yet. When there were surrounding rays with cords and contractures, a nodule was also excluded, because contracture may interfere with the US image. Rays with previous surgery were excluded because of the possibility of scar tissue disturbing the US image.

Procedures

The first observer (R.J.M.v.S.) was trained in the sonographic evaluation of DD patients by the second observer (S.M.) by examining 10 patients with US together. The first observer examined an additional 20 DD patients in between the measurements that were performed together with the second observer, before including patients for the study. The second investigator had already acquired experience with US in DD before the start of this study (> 50 patients) and was initially trained by a clinician with more than 10 years of experience in sonography of DD patients. An US protocol was created by assessing multiple DD patients together before the start of the study. This was done so that, during the study, each nodule was assessed in the same way and the influence of other parameters like US settings, probe direction, pressure, and the amount of US gel used was minimized. The Esaote MyLab 1 device (Genova, Italy) was used, with a 18-mHz probe and the following settings: depth 2 cm, focus 0.5 cm, X-view 1, graymap 2, ambient light 3, dynamic range 8, colorize blue line 3, sharpness 4, and persistence 4.

Transverse and sagittal US images of the selected DD nodule from each participant were obtained and assessed for maximum diameter (width and depth), maximum circumference, and maximum area in both planes (Fig. 1). Width and depth of the nodules were measured using built-in software on the US device. Circumference and area were calculated by drawing a thin line on the outer border of the nodule, using a computer program (Image J).[8]

For the intraobserver reliability, the first observer assessed the selected nodules twice, with a period of at least 1 week, but no more than 2 weeks in between. This period was chosen to limit the possibility of progression, while the first observer was not able to remember the first measurements. For the interobserver agreement, the second observer obtained images once, on the same day as the second measurement of the first observer.

The observers performed their measurements separately so they would not influence each other. The first observer drew the selected nodule on a case record form containing a schematic image of a hand to make sure the same nodule was assessed during all measurements. Furthermore, the participants were instructed not to inform the observers of each other's findings.

Statistical methods

Descriptive statistics were presented by means and SDs (for normally distributed continuous variables).

The intraobserver reliability was calculated using the intraclass correlation coefficient (ICC): 2-way mixed effects model, single measures, absolute agreement. The interobserver reliability was also calculated using the ICC, but with a different model, because with multiple observers, there is also observer variance: 2-way random effects model,

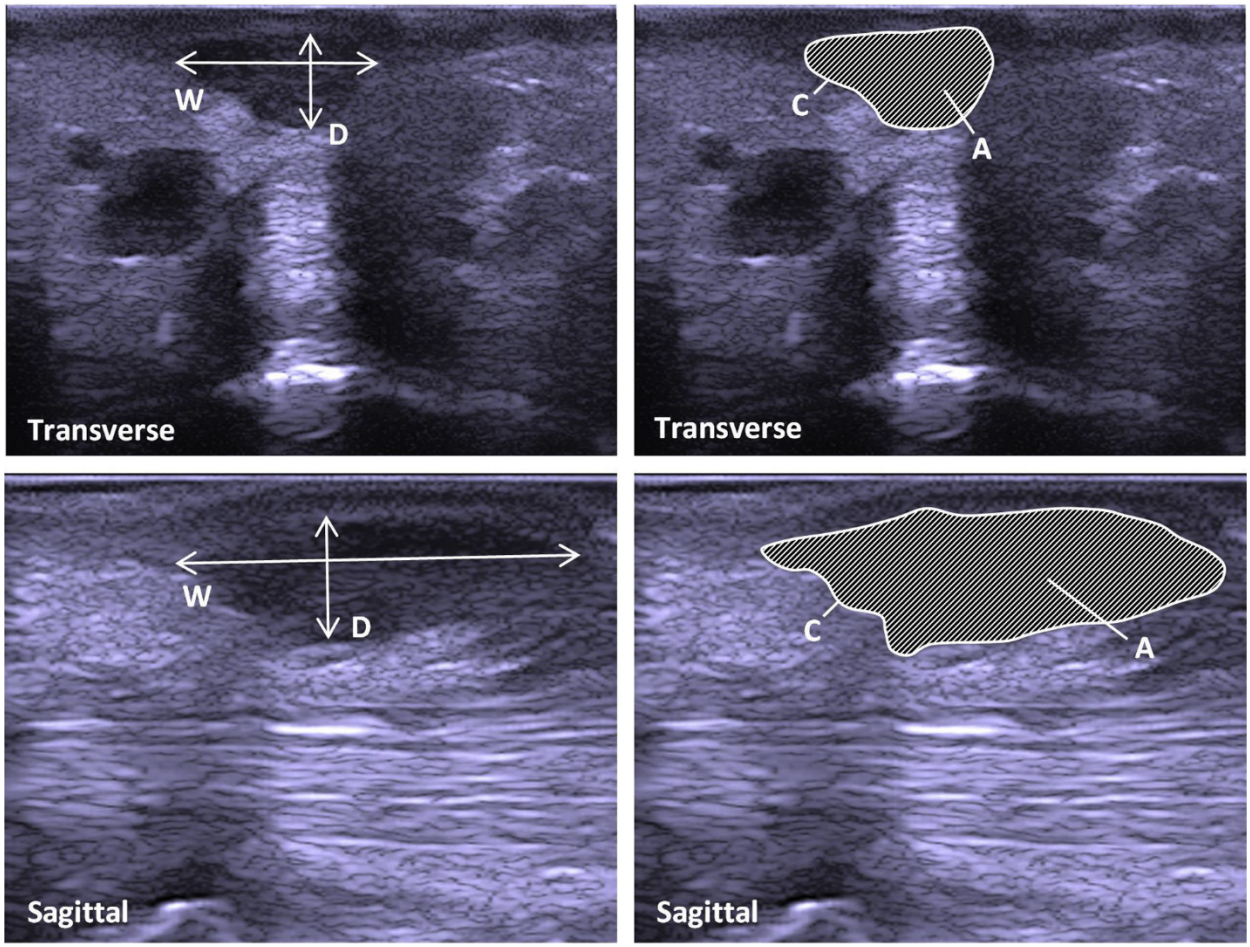


FIGURE 1: US measurements of a nodule in 2 planes. A, area; C, circumference; D, depth; W, width. (Thickness of the lines was exaggerated to improve quality of the images.)

single measures, absolute agreement. Several rating scales for the ICC have been described. Often 0.70 is recommended as a minimum standard for reliability, which is why we defined an ICC of 0.70 or greater as good reliability.[9]

All differences in measurement scores were plotted against the mean scores according to the Bland-Altman method using 95% limits of agreement (1.96*SD).[10]

The SEM was calculated with the following formulas[11]:

- Intraobserver reliability: $\text{SEM} = \sqrt{\sigma^2_{\text{error}}}$
- Interobserver reliability: $\text{SEM} = \sqrt{\sigma^2_{\text{observer}} + \sigma^2_{\text{error}}}$

The variances were obtained through analysis of variance components.

The SEM is a measure of how much the repeated measurements of the outcomes are spread around a true score. Subsequently, the SDC can be calculated, which represents the minimum change that has to be observed to ensure that the change is real and not based on measurement error.

The SDC was calculated using the following formula: $\text{SDC} = 1.96 \times \sqrt{2} \times \text{SEM}$.

RESULTS

A total of 83 patients with early DD nodules were asked to participate, 50 patients gave written informed consent. The mean age was 58.9 years (SD, 8.0) and 29 participants were men (58%). The average size of the different aspects of the nodules was calculated for all 3 measurements together. Overall, and not surprisingly, the average measurements were larger in the sagittal plane than in the transverse plane, except for depth, which showed a comparable result. There was a considerable difference in the average size of all aspects of the nodules between the measurements of observers 1 and 2 (Table 1).

TABLE 1. Average Measurements of Observers 1 and 2: Mean (SD)

	Observer 1.1	Observer 1.2	Observer 2	All 3 Measurements
Transverse				
Width (mm)	8.62 (3.40)	8.15 (2.91)	7.33 (1.93)	8.03 (2.41)
Depth (mm)	3.11 (1.40)	3.13 (1.29)	2.75 (1.00)	3.00 (1.15)
Circumference (mm)	23.11 (8.90)	20.93 (6.73)	19.01 (4.86)	21.02 (6.17)
Area (mm^2)	24.46 (17.24)	22.42 (14.24)	17.35 (8.51)	21.41 (12.27)
Sagittal				
Width (mm)	13.53 (4.13)	13.30 (3.81)	12.61 (3.38)	13.15 (3.13)
Depth (mm)	3.29 (1.31)	3.25 (1.29)	2.79 (0.87)	3.11 (1.06)
Circumference (mm)	35.16 (12.26)	34.51 (10.46)	30.98 (7.43)	33.55 (8.98)
Area (mm^2)	39.81 (24.64)	38.82 (22.56)	31.31 (15.23)	36.64 (19.24)

Reliability and SEM

When looking at the intraobserver reliability, all ICCs were greater than 0.7 except for the ICC of the measurement of nodule width in the sagittal direction. The measurement of nodule area had the highest intraobserver ICC in both directions. When looking at the interobserver reliability, the highest ICCs were obtained for nodule depth, and only depth in the transverse direction exceeded 0.7. Measurement of nodule width in the sagittal direction also had the lowest interobserver ICC.

Overall, the Bland-Altman plots showed that the mean differences of all measured aspects were close to 0. No obvious trends were observed in the scatter around the mean. However, all plots showed several outliers and relatively wide 95% limits of agreement, which were smaller for the intraobserver measurements. The Bland-Altman plots of all measured aspects are shown in Appendix A (available on the *Journal*'s Web site at www.jhandsurg.org).

The SEMs and the SDCs were higher for all measurements performed by different observers, except for depth in the transverse plane. The SEMs and SDCs of all measured aspects were higher in the sagittal plane than in the transverse plane. When comparing the SEMs and SDCs with the average size of nodules, the SEMs and SDCs were relatively smaller in the sagittal plane.

An overview of the study results is given in Table 2.

DISCUSSION

Previous studies have used US for the follow-up of early DD,[5,12] but the reliability of this newly introduced measurement instrument has not been determined. The aim of this study was to determine the reliability and interpretability of US for the measurement of DD nodules when performed by a single observer and when performed by 2 observers.

In total, 50 DD patients with palmar nodules were examined by 2 observers. Overall, the reliability of the measurements performed by a single observer was better than the measurements performed by 2 different observers. When assessing the intraobserver reliability, the ICCs were good (> 0.7) for all measurement directions, except for width in the sagittal plane (ICC, 0.671). Although, clinically, we only selected nodules that were not evidently part of a cord, with US the nodules often appeared cordlike, with longitudinally oriented fibers extending to the proximal and distal border of the US image, the skin, and the underlying tendons. This may have caused a wider spread in the interpretation of the nodule borders and may explain why the ICC of sagittal width was lower than of the other measurements. An example of nodules with borders that were difficult to measure in the sagittal plane is shown in Figure 2B.

The interobserver reliability was lower and had wider confidence intervals. In our experience, both observer-specific factors (ie, probe angle) and patient-specific factors (ie, differences in echogenicity and distinctness of nodule borders) seem to influence the interobserver reliability. The measurement of depth was the only measurement with good agreement, with an ICC in the transverse plane greater than 0.7 and an ICC in the sagittal plane just below 0.7 (ICC, 0.66). However, this was not very surprising because the variation in depth between nodules is expected to be quite small given that they are always situated between the skin and the underlying tendons.

The SEMs of all measured aspects show that the overall deviation of measurements around the true score was relatively small. However, to be able to

TABLE 2. Overview of Study Results

	Width	Depth	Circumference	Area (mm)
Transverse				
Intraobserver				
ICC (95% CI)	0.72 (0.56–0.83)	0.81 (0.69–0.89)	0.80 (0.63–0.76)	0.85 (0.74–0.89)
SEM (mm)	1.65	0.59	3.25	6.10
SDC (mm)	4.57	1.64	9.01	16.91
Interobserver				
ICC (95% CI)	0.49 (0.26–0.68)	0.76 (0.54–0.87)	0.61 (0.39–0.76)	0.57 (0.30–0.74)
SEM: mm	1.79	0.58	3.75	8.02
SDC: mm	4.96	1.61	10.39	22.23
Sagittal				
Intraobserver				
ICC (95% CI)	0.67 (0.48–0.80)	0.79 (0.66–0.88)	0.83 (0.72–0.90)	0.89 (0.81–0.93)
SEM (mm)	2.29	0.60	4.73	8.02
SDC (mm)	6.35	1.66	13.11	22.23
Interobserver				
ICC (95% CI)	0.39 (0.13–0.60)	0.66 (0.38–0.81)	0.57 (0.32–0.74)	0.59 (0.35–0.76)
SEM (mm)	2.83	0.69	6.16	12.71
SDC (mm)	7.84	1.91	17.07	35.23

95% CI, 95% confidence interval.

consider an observed change as a real change instead of measurement error, the SDC has to be calculated. When 2 measurements are compared, 2 sources of measurement error exist because both outcomes are measured with a certain error. The SDCs showed that nodule size has to increase or decrease a large amount before it is certain that the observed change is not based on measurement error. For example, the SDCs showed that nodule size has to increase or decrease by 16.91 (transverse) and 22.23 (sagittal) mm^2, which were 79% and 61% of the average nodule size, respectively, before it is certain that the observed change is not based on measurement error. This indicates that there is a lot of noise around each measurement and, therefore, a change has to exceed this noise before it can be considered as a real change.

All Bland-Altman plots showed that the mean differences were close to 0, which shows that systematic inter- and intraobserver differences were small. However, for every aspect, there were wide 95% limits of agreement and several obvious outliers, but in absence of previous literature on sonographic measurements of DD nodules, it is difficult to interpret this. As mentioned previously, several factors contribute to variability in observations. Because nodules were selected with clinical examination and not based on US features, several nodules were difficult to examine or were not visible at all. Figure 2 shows different scenarios of nodules that may have contributed to the wide limits of agreement and SDCs. This could have been prevented by using stricter inclusion criteria for nodules. However, because no information on reliability of US was available before the start of this study, we were interested in the agreement and accuracy of US for the measurement of palmar nodules in the general DD population. Also, because we had no knowledge of the reliability of US for DD, we blinded the researchers for previous and/or each other's measurements, which is the first step in determining the minimum reliability of a new measurement tool.

Because no previous study has determined the reliability of sonographic measurement of DD nodules, we calculated ICCs, SEMs, and SDCs of different aspects of a nodule (width, depth, circumference, and area). However, in the daily practice, it is often more practical to select 1 aspect and use this to compare measurements throughout a follow-up period. For future studies, we would recommend the measurement of the maximum area by a single observer because the ICCs of area were satisfactory and area reflects the 2-dimensional size of a nodule in

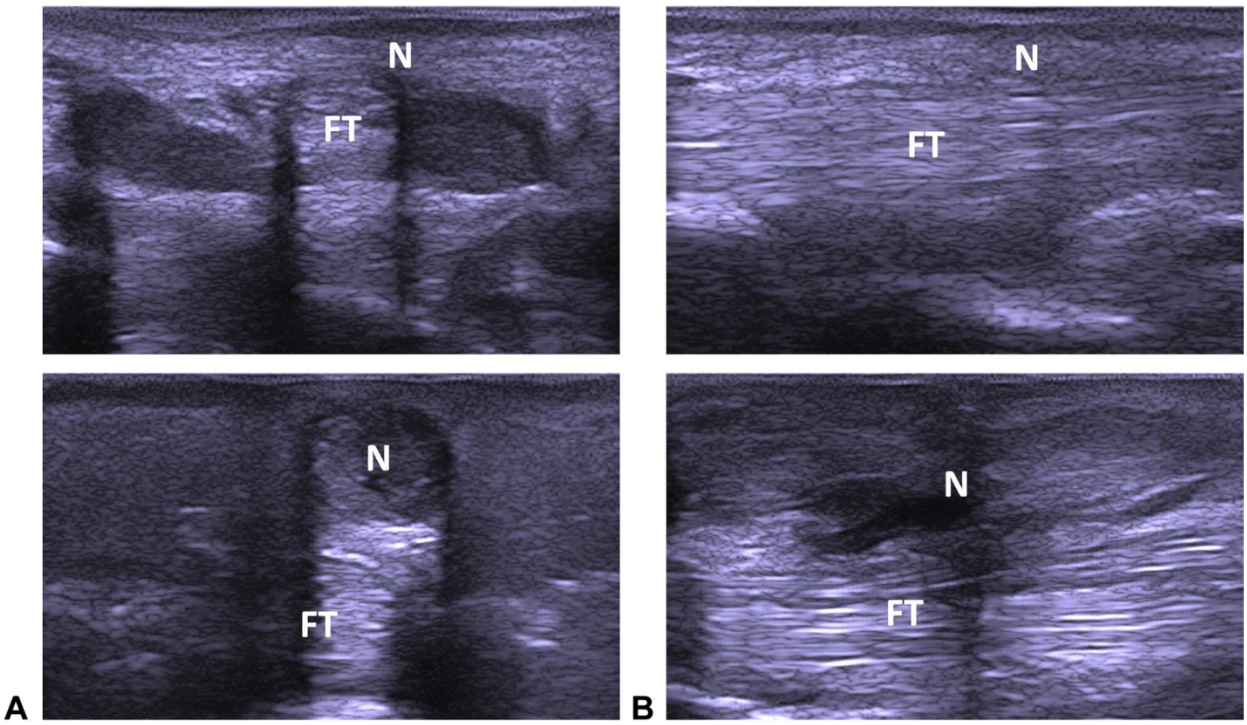


FIGURE 2: Two examples of how inconsistent intra- and interobserver measurements can occur. FT, flexor tendons; N, nodule. **1** **A** Transverse and **B** sagittal images of a nodule that was clear upon physical examination, but not visible with US. **2** Images of a nodule that could very clearly be visualized with US in the **A** transverse plane, but had very unclear borders and mixed echogenicity in the **B** sagittal plane.

1 plane. It is debatable whether a nodule should be measured in the transverse or the sagittal plane. Ideally, when following patients with early DD, nodules should be measured in all 3 planes, because it is unknown whether a nodule progresses equally in each direction. Unfortunately, the frontal plane can only be measured with physical examination, which has its limitations. When choosing a single plane, we would advise selecting the sagittal plane because measurement of area in this direction is somewhat more precise. Also, most previous studies that show images of patients with DD describe sonographic measurement in the sagittal plane, which makes results more comparable.[5,6,12]

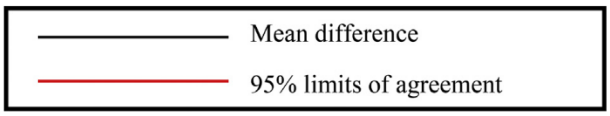
Because US is not regularly used for patients with DD, it is not known what the learning curve is for the visualization of nodules. In our experience, because DD is a palpable disease, it is relatively easy to assess with US when the examiner has a thorough knowledge of the anatomy of the hand. After assessing 30 patients, both examiners felt confident enough to measure different types of nodules and start including patients for research purposes. It could be that reliability of multiple measurements still increases after these 30 patients, which may be seen as a limitation of this study. Another limitation could be that the short period (1−2 weeks) between the measurements of observer 1, which was chosen to reduce the risk of progression in the patients, led to recall bias.

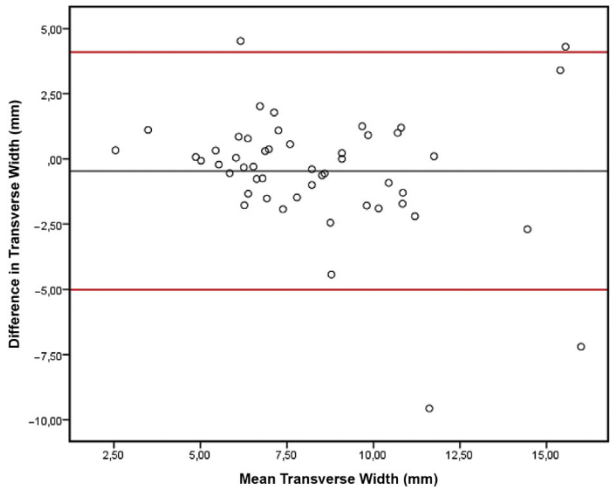
The SEMs and SDCs found in our study were larger than we expected. As mentioned previously, we could have used stricter criteria for the inclusion of nodules. If only the nodules that are clearly visible on US were included, this would probably have led to higher ICCs and lower SEMs and SDCs. This may be seen as a third limitation of our study. However, narrowing the inclusion criteria would also have compromised the generalizability of our results. Future studies using US for palmar DD nodules may choose to use stricter inclusion and exclusion criteria. Also, researchers may choose to use previous images and compare these with new measurements. Alternatively, researchers may decide to measure nodules multiple times during each measurement and use the average measurements to calculate the SEMs and SDCs, which are likely to be smaller this way.

REFERENCES

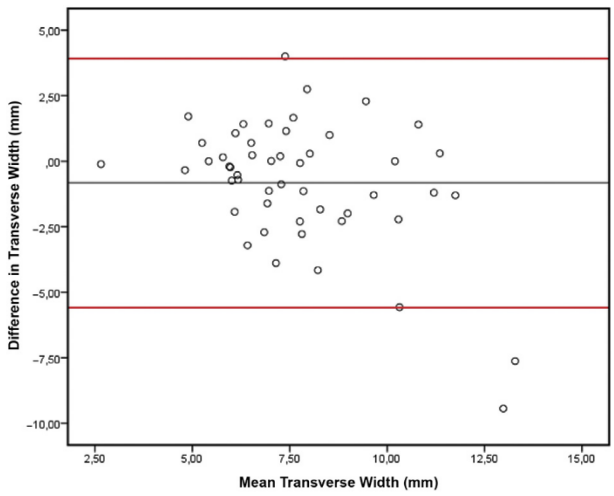
1. Werker PMN, Degreef I. Alternative and adjunctive treatments for Dupuytren disease. *Hand Clin*. 2018;34(3):367–375.
2. Degreef I, De Smet L. Dupuytren disease: on our way to a cure? *Acta Orthop Belg*. 2013;79(3):243–249.
3. Nanchahal J, Ball C, Davidson D, et al. Anti-tumour necrosis factor therapy for Dupuytren's disease: a randomised dose response proof of concept phase 2a clinical trial. *EBioMedicine*. 2018;33:282–288.
4. Broekstra DC, Lanting R, Werker PM, van den Heuvel ER. Intra- and inter-observer agreement on diagnosis of Dupuytren disease, measurements of severity of contracture, and disease extent. *Man Ther*. 2015;20(4):580–586.
5. Yin CY, Yu HM, Wang JP, Huang YC, Huang TF, Chang MC. Long-term follow-up of Dupuytren disease after injection of triamcinolone acetonide in Chinese patients in Taiwan. *J Hand Surg Eur Vol*. 2017;42(7):678–682.
6. Creteur V, Madani A, Gosset N. Ultrasound imaging of Dupuytren's contracture. *J Radiol*. 2010;91(6):687–691.
7. Read JW, Conolly WB, Lanzetta M, Spielman S, Snodgrass D, Korber JS. Diagnostic ultrasound of the hand and wrist. *J Hand Surg Am*. 1996;21(6):1004–1010.
8. Abramoff MD, Magalhaes PJ, Ram SJ. Image processing with ImageJ. *Biophotonics Int*. 2004;11(7):36–42.
9. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.
10. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician*. 1983;32(3):307–317.
11. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Reliability. *Measurement in Medicine: A Practical Guide*. New York: Cambridge University Press; 2011:96–149.
12. Christie WS, Puhl AA, Lucaciu OC. Cross-frictional therapy and stretching for the treatment of palmar adhesions due to Dupuytren's contracture: a prospective case study. *Man Ther*. 2012;17(5):479–482.

APPENDIX A. BLAND-ALTMAN PLOTS

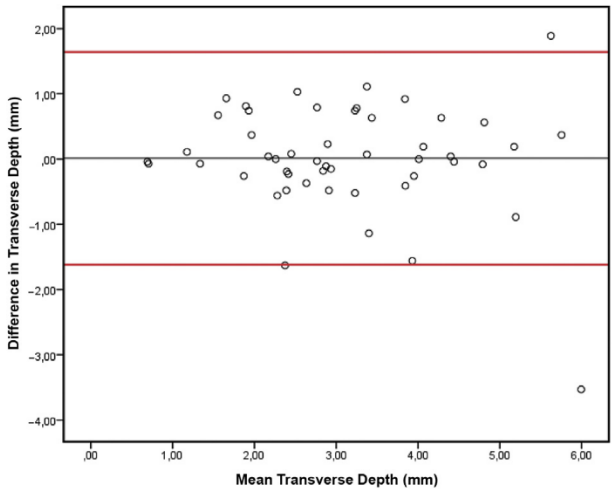




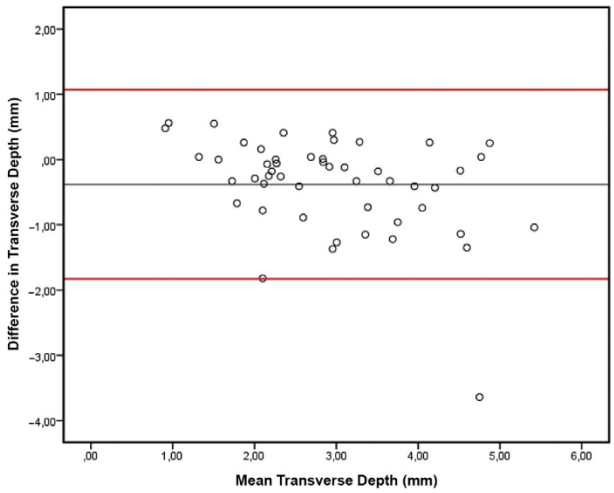
TRANSVERSE WIDTH—INTRAOBSERVER



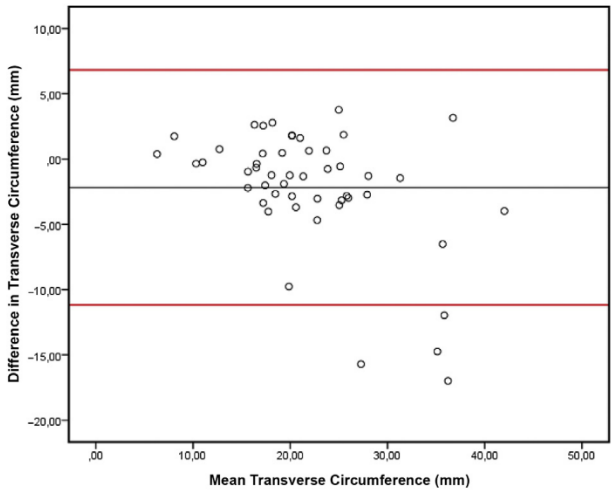
TRANSVERSE WIDTH—INTEROBSERVER



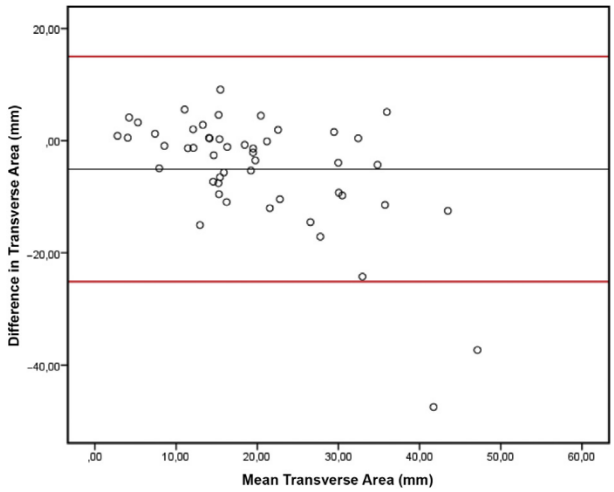
TRANSVERSE DEPTH—INTRAOBSERVER



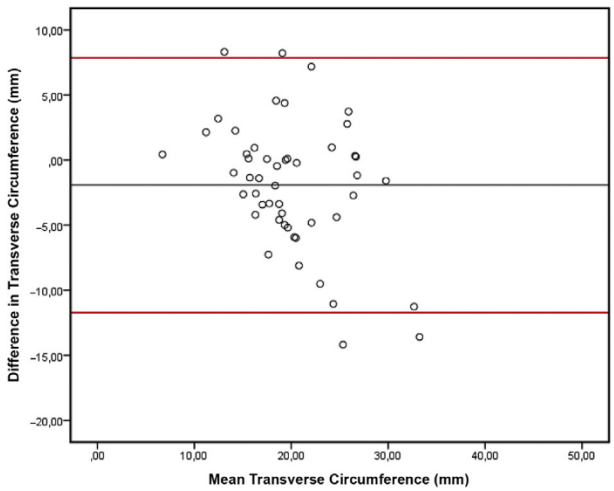
TRANSVERSE DEPTH—INTEROBSERVER



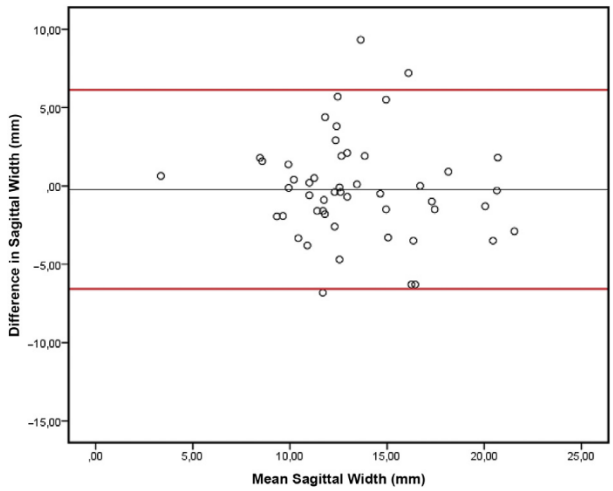
TRANSVERSE CIRCUMFERENCE—INTRAOBSERVER



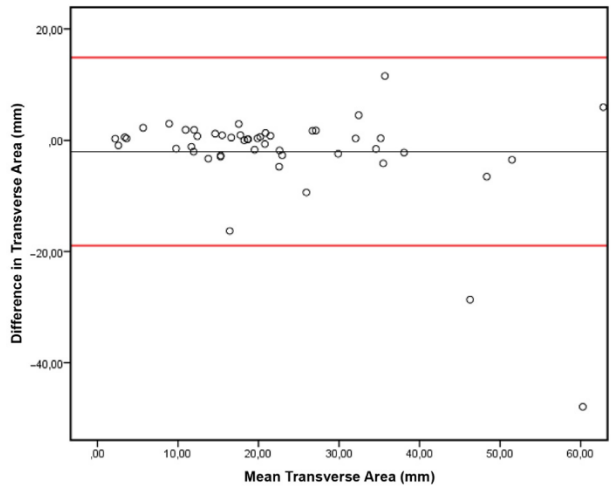
TRANSVERSE AREA—INTEROBSERVER



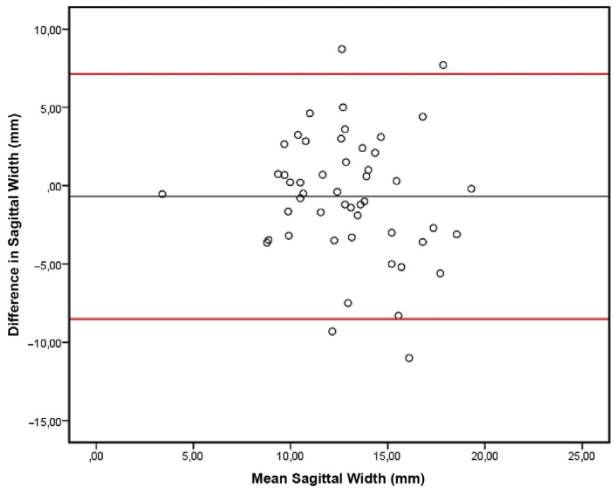
TRANSVERSE CIRCUMFERENCE—INTEROBSERVER



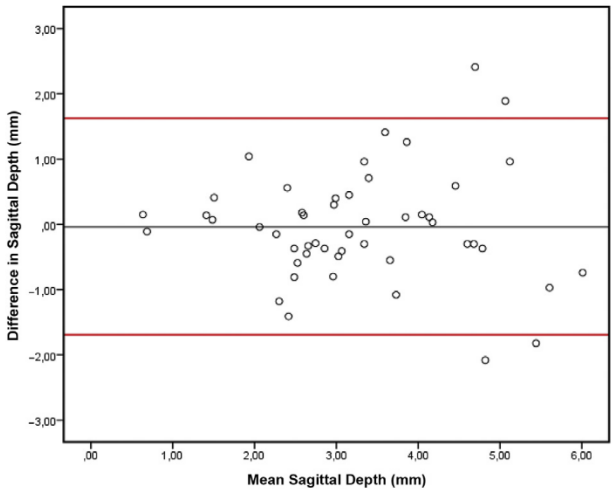
SAGITTAL WIDTH—INTRAOBSERVER



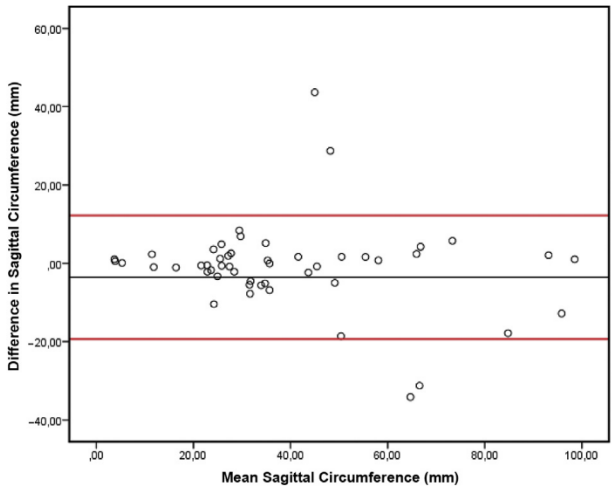
TRANSVERSE AREA—INTRAOBSERVER



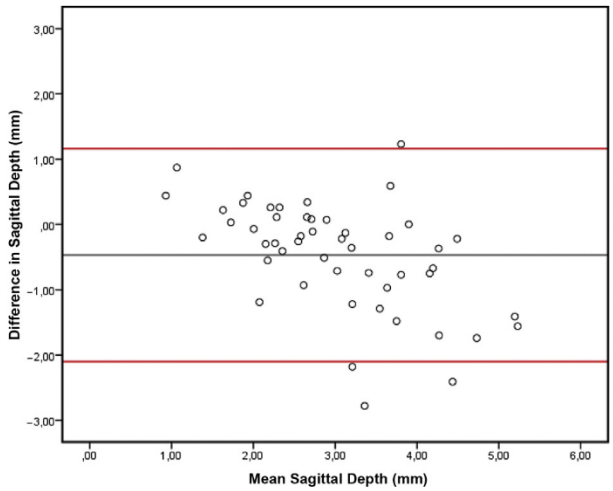
SAGITTAL WIDTH—INTEROBSERVER



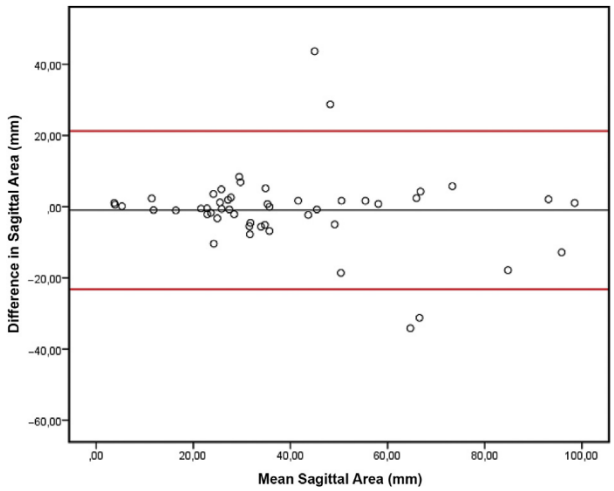
SAGITTAL DEPTH—INTRAOBSERVER



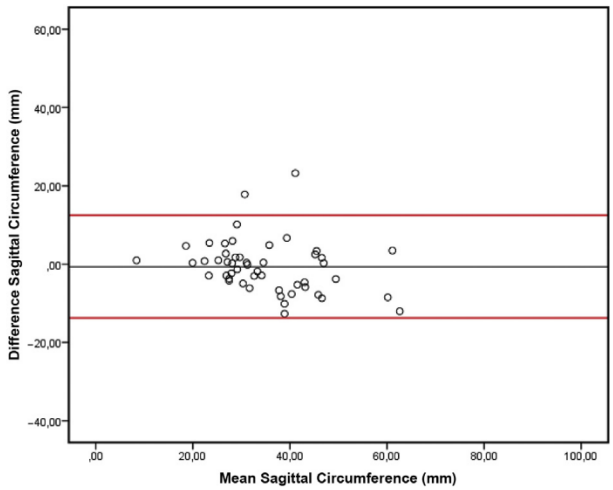
SAGITTAL CIRCUMFERENCE—INTRAOBSERVER



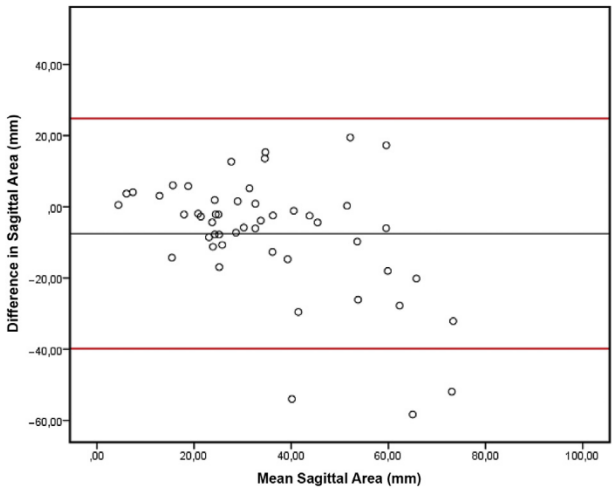
SAGITTAL DEPTH—INTEROBSERVER



SAGITTAL AREA—INTRAOBSERVER



SAGITTAL CIRCUMFERENCE—INTRAOBSERVER



SAGITTAL AREA—INTEROBSERVER