Original paper

Accuracy and reliability of indices for assessing proximal interphalangeal joint position in hammer toes

Ewa Puszczałowska-Lizis[1], Bartosz Białas[2], Dominik Szymański[3]

[1]University of Rzeszów, *Collegium Medicum*, Faculty of Health Sciences and Psychology, Rzeszów, Poland
[2]Non Public Physiotherapy Office in Rzeszów, Rzeszów, Poland
[3]Medical Rehabilitation and Osteopathy Clinic "REHApunkt", Warsaw, Poland

Medical Studies
DOI: https://doi.org/10.5114/ms.2026.160442

Key words: telemedicine, test-retest, flexion at the interphalangeal joint, hammer toe.

Abstract

Introduction: Hammer toes are among the most common toe deformities.
Aim of the research: This study evaluated the accuracy and measurement reliability of the six indices proposed by the authors for assessing deformities typical of hammer toes.
Material and methods: We examined 30 persons with hammer toes, aged 65–74 years. Measurements were compared using a GIMA anthropometric tape (Gima S.p.A., Gessate (MI), Italy), Beerendonk HOSSA caliper (Hossa Int., Sialkot, Pakistan), Xiaomi Redmi 5 smartphone camera (Xiaomi Inc., Beijing, China), and the GIMP application. The results were analysed using the Wilcoxon test and factor analysis. A test-retest measures design was utilised, with the intraclass correlation coefficient (ICC) used to determine levels of reliability.
Results: The index of height of the proximal interphalangeal joint (H1), and the index determining the difference in the height of the examined joint compared to interphalangeal joint of the big toe (H3), achieved the maximum factor loading values (–0.831 and 0.820, respectively). The values of intraclass correlation coefficients for all assessed indices exceeded 0.90, with the highest values recorded for the α angle (ICC = 0.99) and the H1 index (ICC = 0.98).
Conclusions: Proposed indices for testing the position of the proximal interphalangeal joint in hammer toes meet the accuracy criterion, with the H1 index being the most accurate for testing the absolute height of this joint, and the H3 index for testing the relative height. All proposed indices meet the criterion of reliability of measurement results, and therefore they can be used in clinical studies to assess the position of the proximal interphalangeal joint in hammer toes.

Introduction

Hammer toes are characterised by a fixed flexion of the proximal interphalangeal joint and loss of contact between the toe and the ground [1, 2]. The disorder can vary in severity. Initially, the mobility of the joint may be preserved, but over time there are structural changes that limit its mobility. In mild cases, the changes do not affect functional performance and are more of an aesthetic problem. Often, however, they significantly impede daily activities and cause persistent discomfort [3, 4].

The definition of hammer toes is a debatable topic, although clinicians generally agree that hammer toes result from flexion at the proximal interphalangeal joint [5]. However, there is a discrepancy in the description of the other joints. This is because some researchers indicate that there is extension at the metatarsophalangeal and distal interphalangeal joints of the affected toe, while others omit these joints from the definition altogether [6, 7]. Correctly defining hammer toes is important for indicating the difference between other toe deformities and implementing appropriate treatment [1]. For the purposes of this study, it was assumed that hammer toe is a deformity of the toes characterised by a correctable or uncorrectable flexion position at the proximal interphalangeal joint and preserved contact of the distal phalanx with the ground in the standing position. According to the available literature, hammer toe is most often found in the longest of the toes II-V, with the vast majority being toe II [6].

Currently, there is a lack of objective and readily available methods for studying hammer toes. A few studies focus on the mobility of the first ray of the foot, particularly the metatarsophalangeal joint, using a goniometer [8]. In clinical practice and research, subjective visual assessment, "feel" of passive motion testing at the proximal interphalangeal joint of the foot, 3D scanners, or invasive radiological methods are used to evaluate toes II-V [8]. In view of the above, the authors believe that the implementation of objective and readily available assessment techniques will, from a scientific point of view, facilitate research into the problems of hammer toes, and

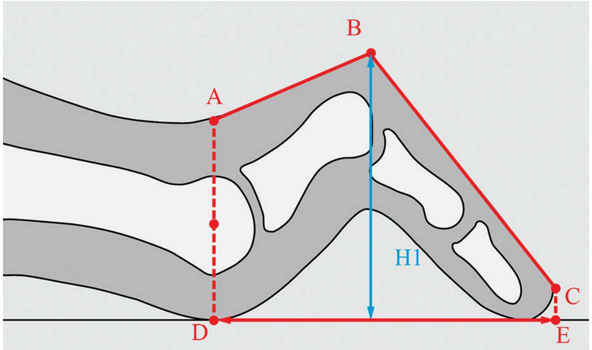

Figure 1. Manner of determining points and sections required to calculate the tested indicators

from a clinical point of view, improve diagnostic processes.

Aim of the research

The aim of the study was to evaluate the accuracy and measurement reliability of the original indices proposed by the authors for assessing deformities typical of hammer toes.

Material and methods

The research included 30 people (15 W; 15 M) in the 65–74 age range ($\bar{x}$ = 68.50 ±4.16 years), who met the following inclusion criteria: age 65–74 years (youngest-old), the presence of at least one hammer toe (taking into account toes II–IV), the absence of hammer toe lesions within the big toe, participation in 1st and 2nd measurement, wearing the same footwear between 1st and 2nd measurement, and providing written informed consent to participate in the study.

Exclusion criteria were as follows: current treatment of hammertoes, lower limb oedema, a history of degenerative or inflammatory joint disease, any neurological disorders, or recent foot surgery, a history of surgery on toes II–V, identified foot deformities other than hammer toes to an extent that affects the size of the deformity under study, recent (past 3 months) lower limb injury, difficulty maintaining a standing position, and lack of consent to participate in the study.

A total of 60 hammertoes were found in all subjects, including 30 in the 2nd toe, 12 in the 3rd toe, and 18 in the 4th toe.

Measurements of each subject's toes were taken using the following testing tools:

1. Anthropometric tape GIMA (Gima S.p.A., Gessate (MI), Italy). Using an anthropometric tape, the following measurements were taken:
   - length of the proximal phalanx – segment |AB|,
   - length of the middle and distal phalanges – segment |BC|,
   - length from the head of the metatarsal bone to the end of the distal phalanx – segment |DE| (Figure 1).
2. Caliper Beerendonk HOSSA (Hossa Int., Sialkot, Pakistan). Using a caliper depth gauge, the following measurements were taken:
   - height of the proximal interphalangeal joint from the ground – H1 (Figure 1),
   - height of the interphalangeal joint of the big toe of the same foot.

   After performing the above measurements, the following indices were determined:
   - H1 index – determines the height of the proximal interphalangeal joint – represents the value of the height measurement from the highest point within the proximal interphalangeal joint to the ground, in mm (Figure 1).
   - H3 index – determines the difference in the height of the examined joint compared to interphalangeal joint of the big toe – the difference in the value of the index H1 and the height H2 (meaning the value of the height measurement from the highest point within the interphalangeal joint of the toe to the ground), in mm: H3 = H1 – H2.
   - W1 index – this is the quotient of the segment |DE|, measured from the head of the metatarsal bone to the end of the distal phalanx, by the sum of the segments |AB| and |BC|, denoting the total length of the middle and distal phalanges (Figure 1), in mm: W1 = |DE|/(|AB| + |BC|).
   - W2 index – this is the quotient of the height from the highest point within the proximal interphalangeal joint to the ground (H1) and the total length of the middle and distal phalanges (|BC|), in mm (Figure 1): W2 = H1/|BC|.
3. Xiaomi Redmi 5 smartphone camera (Xiaomi Inc., Beijing, China). Testing joint ranges of motion using photos taken with a smartphone is a popular and reliable method of testing [9–11]. Photos were taken in lateral projection, recording the toe in the sagittal plane. The obtained images were used to determine the following indices: α angle and WF1 index. The GNU Image Manipulation Program (GIMP) software was used for this purpose.
   - α angle – the angle of flexion at the proximal interphalangeal joint of the toe in the foot, in degrees. When developing this index, we took into account the proposal of Zhao's *et al.* [11] for determining the range of flexion in the finger joints. Before the photo was taken, the highest point of the proximal interphalangeal joint was marked on the subject's skin, which served as a reference point for plotting the α angle, contained between lines tangent to the dorsal surface of the proximal phalanx and middle phalanx (Figure 2).
   - WF1 index – this was determined on a similar basis as the W1 index, but using a camera and

a computer application. Before the photo was taken, the highest point of the proximal interphalangeal joint was marked on the patient's skin, from which sections representing the length of the proximal phalanx and the common length of the middle and distal phalanges were determined using the GIMP software, analogous to the W1 index.

To assess the intra-rater reliability – i.e. the extent to which the results of two different tests, performed by the same researcher, using the same measurement tools, under identical conditions, are free of random error – measurements of individual features were made 2 times by the same person, 1 week apart [12].

The normal distribution of the values was verified with the Shapiro-Wilk test. To compare the results obtained in 1st (test) and 2nd (retest) examination the Wilcoxon test was used. Verification of the accuracy of the selection of indices for assessing the height of the proximal interphalangeal joint in the case of hammer toes was carried out using factor analysis [13]. This method was performed taking into account the mean values obtained for each index in two measurements.

The number of factors was determined on the basis of a scree plot, taking into account the Kaiser criterion that the eigenvalue corresponding to the common factor cannot be less than 1.0, i.e. the variance "transferred" by a factor cannot be less than the variance corresponding to a single variable. An orthogonal rotation of the factors was conducted using the normalised Equamax method [14]. In the interpretation of the results, loadings with a value greater than 0.7 were considered significant and therefore to show a high degree of interdependence within the factor [13].

The reliability of the proposed indices was tested using the test-retest method by calculating the intraclass correlation coefficient (ICC), with interpretation according to Koo and Li [15]: below 0.50 poor; 0.50–0.75 moderate; 0.75–0.90 good; above 0.90 excellent. For each of the obtained ICCs, the standard error of measurement (SEM) and the minimum detectable change (MDC) for the 95% confidence interval were calculated [14].

Statistical analysis

The significance level was set at $p < 0.05$. Statistical computations were performed using Statistica v13.3 and SPSS Statistics v24.0.

Results

There were no statistically significant differences in the values of indices obtained in both measurements (Table 1).

Table 2 contains values defining the variance explained by particular factors. The greater the eigenvalue (variance) of a factor, the greater the "information load" of the input variables the factor contains. The first factor accounts for the highest percentage of variance. The second explains most of what remains to be explained, etc. The percentage of total variance explained by a given factor was calculated by dividing the factor's eigenvalue by the number of variables introduced into the study and then multiplying the obtained result by 100. Calculations performed in this way revealed the appearance of two factors explaining approximately 73% of the information contained in the input variables, as demonstrated by the value of the cumulative percentage of variance.

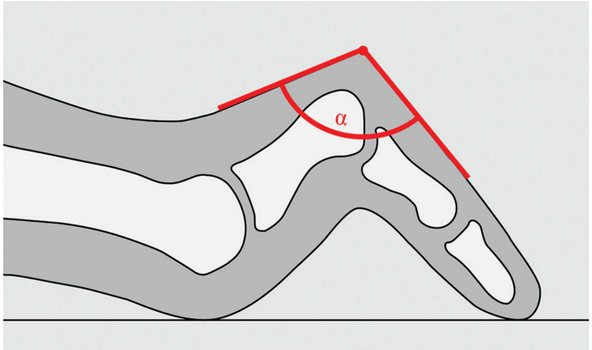

Figure 2. Manner of determining the α angle

Based on eigenvalues (Table 2) and the scree plot (Figure 3), two factors were determined whose eigenvalue exceeded 1 and therefore explained more than one of the analysed variables.

Data in Figure 3 show that the second factor is the last factor with an eigenvalue greater than one. This allowed us to conclude that the scree begins with the eigenvalue of the second factor; therefore, two factors were adopted for further analysis.

Table 3 contains the values of factor loadings for the analysed variables after rotation. Analysis of the structure of the studied variables indicates that the common variance was explained to the greatest extent (44%) by the first factor, which is identified by the following indices: height of the proximal interphalangeal joint H1, height relative to the length of the two distal phalanges W2, α angle, and the ratio of absolute length to relative length on the photo WF1. The H1 index (–0.831), and α angle (0.789) were most responsible for the variance in factor one. This factor was called "absolute pond height" because it contains a high factor loading value for the H1 index.

The second factor was isolated on the basis of two variables and explained 29% of the common variance. This factor is related to the variables: the difference in the height of the joints H3 (0.820) and the W1 index (0.748). This factor was called the "relative pond height".

Figure 4 shows the values of factor loadings for each index in the case of both factors.

Table 4 contains the values of intraclass correlation coefficients between the 1st and 2nd measurement

Table 1. Comparison of the values of indices obtained in both measurements

Exam	$\bar{x}$ ± SD	Max.–Min.	Q_{25}	Me	Q_{75}	*Z*	*P*-value
H1 index [mm]							
Test	2.47 ±0.34	3.21–1.89	2.20	2.47	2.79	1.40	0.150
Retest	2.48 ±0.34	3.15–1.87	2.20	2.40	2.80		
H3 index [mm]							
Test	0.11 ±0.29	0.64–(–0.55)	–0.14	0.13	0.35	0.75	0.454
Retest	0.11 ±0.29	0.70–(–0.55)	–0.14	0.13	0.35		
W1 index [mm]							
Test	0.56 ±0.07	0.80–0.38	0.52	0.56	0.60	0.31	0.754
Retest	0.56 ±0.07	0.82–0.41	0.51	0.56	0.60		
W2 index [mm]							
Test	0.84 ±0.08	1.07–0.66	0.78	0.83	0.89	0.39	0.694
Retest	0.84 ±0.09	1.08–0.68	0.78	0.83	0.90		
α angle [°]							
Test	139.40 ±14.38	161.00–110.00	129.50	143.50	149.00	0.34	0.735
Retest	139.38 ±14.23	162.00–110.00	128.00	144.00	150.00		
WF1 index [mm]							
Test	0.79 ±0.06	0.89–0.60	0.75	0.80	0.84	0.24	0.812
Retest	0.79 ±0.06	0.88–0.56	0.77	0.79	0.83		

$\bar{x}$ – arithmetic mean value, SD – standard deviation, Max – maximum value, Min – minimum value, Q25 – lower quartile, Me – median, Q75 – upper quartile, Z – value of the Wilcoxon matched pairs test statistics, p – probability value.

Table 2. Eigenvalue and the amount of variance explained by particular factors before selection and without rotation

Factor	Eigenvalue	Cumulative eigenvalue	% of total variance	Cumulative % variance
1	**2.67**	2.67	44.58	44.58
2	**1.68**	4.35	27.92	72.50
3	0.76	5.11	12.66	85.16
4	0.52	5.63	8.70	93.86
5	0.22	5.86	3.73	97.59
6	0.14	6.00	2.41	100.00

Eigenvalues are given in bold.

for particular indices. The values of these coefficients for all assessed indices exceeded 0.90, with the highest values recorded for the α angle (ICC = 0.99) and the H1 index (ICC = 0.98). It is worth mentioning that for the α angle the highest values of the standard error of measurement (SEM = 1.36) and the minimal detectable change (MDC = 3.76) were recorded.

Discussion

The indices proposed by the authors for testing the position of the proximal interphalangeal joint in hammer toes meet the accuracy criterion, with the H1 index being the most accurate for testing the absolute height of this joint, and the H3 index for testing the relative height. Perhaps this is due to the greater simplicity in determining the measurement points to determine these indices. It is worth mentioning that the angle α, which determines the flexion of the toe at the proximal interphalangeal joint, was also largely responsible for the variance of factor one. This result is reasonable and consistent with the assumptions considering the inclusion of only the proximal interphalangeal joint in the definition of hammer toe. The lower value of factor loadings relative to the H1 index can be explained by a greater risk of measure-

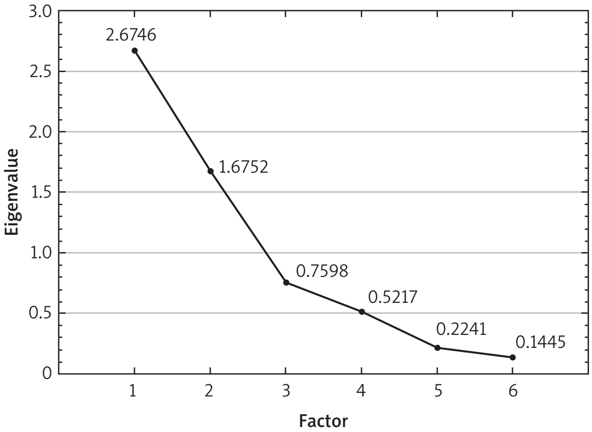

Figure 3. Graph of the scree

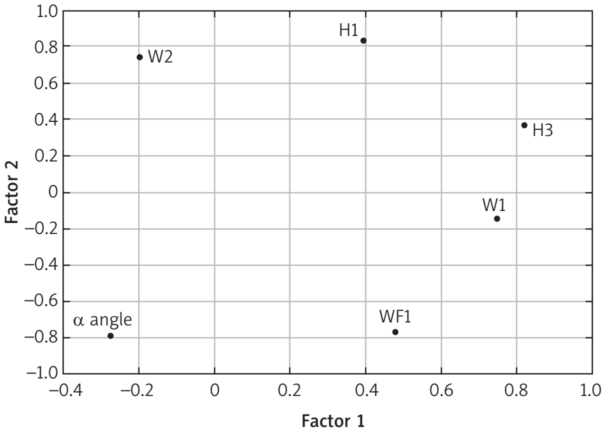

Figure 4. Scatterplot for factor loadings of both factors

Table 3. Values of factor loadings after rotation using the normalised Equamax method

Factor	Factor 1 "absolute pond height"	Factor 2 "relative pond height"
H1	**–0.831**	0.394
H3	–0.371	**0.820**
W1	0.144	**0.748**
W2	**–0.747**	–0.196
α angle	**0.789**	–0.273
WF1	**0.770**	0.478
Eigenvalue	2.621	1.728
Contribution of variance	0.437	0.288
Total variability 73%	44%	29%

The load factor values of over 0.7 (i.e. significant primary features entering the composition of individual main factors) are given in bold.

Table 4. Test-retest reliability of the examined indices

Variable	ICC	SEM	MDC (95% CI)	*R*
H1	0.98	0.04	0.12	0.98
H3	0.96	0.06	0.16	0.96
W1	0.97	0.01	0.03	0.94
W2	0.90	0.03	0.08	0.90
α angle	0.99	1.36	3.76	0.98
WF1	0.92	0.02	0.04	0.93

ICC – intraclass correlation coefficient, SEM – standard error of measurement, MDC – minimum detectable change, R – Spearman's rank correlation coefficient.

ment error, which may be due to technical issues, or the quality of the camera used for the measurements. Our study of hammer toes is pioneering, and the results, in terms of accuracy of flexion angle measurements based on photography, are in line with studies by other authors on other joints, including the wrist and fingers of the hand [11, 16].

Our study showed that all the proposed indices meet the test-retest reliability criterion and can therefore be used in clinical studies to assess the position of the proximal interphalangeal joint in hammer toes. These indices can be an alternative to more complicated, costly, and time-consuming methods, such as X-ray or CT imaging studies. It is noteworthy that the highest reproducibility, i.e. the degree to which the results of two different tests, performed by the same person, using the same measurement tools, were free of random error, was characterised by the H1 index. This is justified because, unlike the others, only one target measurement is required to determine this index. A high value for the test-retest correlation coefficient was also presented by the α angle. However, in addition to excellent reliability, high values of standard error of measurement and minimum detectable difference of measurements were also reported.

Due to the small number of studies with similar subjects, it is difficult to relate the results to those of other authors. Among them are the studies by Jonson *et al.* [17] on the reliability of goniometric measurements in the interphalangeal joints of the hand fingers based on photographs, in which the authors obtained high values of correlation coefficients. In turn, Otter *et al.* [18] investigated the intra-rater reliability together with the concurrent validity of a smartphone goniometric application compared with a traditional goniometer when measuring both joint position and passive motion of the first metatarsophalangeal joint dorsiflexion. The authors found inter-rater reliability lower than in our research, which in both cases was at a good level. Kwon *et al.* [8] proved the relevance and reliability of 3D scanner and computed tomography as methods for measuring the alignment of hammer toes. However, these methods relied on the angular position of the metatarsophalangeal joint, which is not always involved in the deformity and in most cases is not considered in the definition of the condition [1, 7]. An alternative way of assessing joint flexion angles was given by Macionis [20], comparing the diagrammatic goniometry method with a standard goniometer. The method consisted of applying two small, intersecting pieces of paper to a joint and then tracing a line where they intersect. The angle, which was the result of the goniometric measurement, was then measured on the sheet of paper with the marked line. The intra-rater reliability of this tool was shown to be almost perfect for experienced testers [19], and the inter-rater reliability comparable to a standard goniometer for inexperienced testers [20]. It is worth noting that the measurement reliability in this case was for the fingers of the hand, so it can be assumed that due to the complicated measurement procedure, the result of the reliability analysis for the toes could be weaker.

In our study, the results for the W1 and WF1 indices differed. Although both indices aim to measure hammer toes in the same way, the method of determining the absolute length of the toe is different for them. For the first index, the metatarsophalangeal joint is marked on the sole side, while for the index determined from a photograph, the joint is previously marked on the patient's skin on the dorsal side. For this reason, these indices should not be used interchangeably or compared with each other. In addition, when determining the location of the joint on the sole side, it is possible to falsify the value of the index by displacement of the fatty tissue due to caudal sliding of the heads of the metatarsal bones [21].

The smartphone camera and GNU Image Manipulation Program (GIMP) used for the study proved sufficient to demonstrate the relevance and reliability of the indices studied. Studies on the effectiveness of measuring other parts of the lower extremities also confirm the effectiveness of this application and the lack of significant differences compared to the results of tests using other applications [22]. The possibility of using this software in assessing the position of the proximal interphalangeal joint offers the possibility of further research and improving the diagnostic process of hammer toes. The authors express their belief that it is possible to automate the measurement of the indices studied. Such a solution would reduce the time required to obtain the final value of the α angle and the WF1 index, and furthermore would enable the diagnosis of deformities in telemedicine.

In our study, only toes II-IV were considered, which may be a limitation. Therefore, it is recommended that further studies be performed on the possibility of using the proposed indices to evaluate other toes.

Conclusions

The proposed indices for testing the position of the proximal interphalangeal joint in hammer toes meet the accuracy criterion, with the H1 index being the most accurate for testing the absolute height of this joint and the H3 index for testing the relative height.

All proposed indices meet the criterion of reliability of measurement results, and therefore, after verifying the other criteria determining the degree of excellence of research tools, they can be used in clinical studies to assess the position of the proximal interphalangeal joint in hammer toes.

Funding

No external funding.

Ethical approval

The study protocol was approved by the Bioethics Review Committee of the University of Rzeszow (No. 3/04/2020).

Conflict of interest

The authors declare no conflict of interest.

References

1. Harrasser N, Toepfer A. Hammer-, Krallen-, Klauenzehen und Schneiderballen. Was Sie als Hausarzt wissen sollten. MMW Fortschritte der Medizin. 2015; 157(11): 43-45.
2. Goransson M, Constant D. Hammertoe. StatPearls, 2022.
3. Golightly YM, Hannan MT, Dufour AB, Jordan JM. Racial differences in foot disorders and foot type. Arthritis Care Res. 2012; 64(11): 1756-1759.
4. Şayli U, Altunok EÇ, Güven M, Akman B, Biros J, Şayli A. Prevalence estimation and familial tendency of common forefoot deformities in Turkey: a survey of 2662 adults. Acta Orthop Traumatol. 2018; 52(3): 167-173.
5. Konkel KF, Sover ER, Menger AG, Halberg JM. Hammer toe correction using an absorbable pin. Foot Ankle Int. 2011; 32(10): 973-978.
6. Coughlin MJ, Mann RA, Haskell A. Surgery of the Foot and Ankle. Mosby-Elsevier Philadelphia, 2007.
7. Schrier JC, Louwerens JW, Verheyen CC. Opinions on lesser toe deformities among Dutch orthopaedic departments. Foot Ankle Int. 2007; 28(12): 1265-1270.
8. Kwon OY, Tuttle LJ, Commean PK, Mueller MJ. Reliability and validity of measures of hammer toe deformity angle and tibial torsion. Foot. 2009; 19(3): 149-155.
9. Chen J, Xian Zhang A, Jia Qian S, Jing Wang Y. Measurement of finger joint motion after flexor tendon repair: smartphone photography compared with traditional goniometry. J Hand Surg Eur. 2021; 46(8): 825-829.
10. Keogh JWL, Cox A, Anderson S, Liew B, Olsen A, Schram B, Furness J. Reliability and validity of clinically accessible smartphone applications to measure joint range of motion: a systematic review. PLoS One. 2019; 14(5): e0215806.
11. Zhao JZ, Blazar PE, Mora AN, Earp BE. Range of motion measurements of the fingers via smartphone photography. Hand. 2020; 15(5): 679-685.
12. Puszczałowska-Lizis E. Assessment of the measurement reliability of plantography indexes. Prz Med Uniw Rzesz Inst Leków. 2010; 8(2): 176-181.

13. Puszczałowska-Lizis E. Verification of the correctness in index choice in foot structure evaluation on the basis of research into the feet of women aged 20 to 27. Med Rehabil. 2011; 15(2): 2-14.
14. Górniak J. Factor analysis and analysis of principal components. ASK. 1998; 7: 83-102.
15. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016; 15(2): 155-163.
16. Wagner ER, Conti Mica M, Shin AY. Smartphone photography utilized to measure wrist range of motion. J Hand Surg Eur. 2018; 43(2): 187-192.
17. Johnson D, Barradas R, Newington L. Inter-rater and intrarater reliability of finger goniometry measured from screenshots taken via video consultation. J Hand Surg Eur. 2023; 48(5): 459-465.
18. Otter SJ, Agalliu B, Baer N, Hales G, Harvey K, James K, Keating R, McConnell W, Nelson R, Qureshi S, Ryan S, John AS, WaddingtoN H, Warren K, Wong D. The reliability of a smartphone goniometer application compared with a traditional goniometer for measuring first metatarsophalangeal joint dorsiflexion. J Foot Ankle Res. 2015; 8: 30.
19. Macionis V. Is diagrammatic goniometry feasible for finger ROM evaluation and self-evaluation? Clin Orthop Relat Res. 2013; 471(6): 1894-1903.
20. Macionis V. Reliability of the standard goniometry and diagrammatic recording of finger joint angles: a comparative study with healthy subjects and non-professional raters. BMC Musculoskelet Disord. 2013; 14: 17.
21. Bus SA, Maas M, Cavanagh PR, Michels RP, Levi M. Plantar fat-pad displacement in neuropathic diabetic patients with toe deformity: a magnetic resonance imaging study. Diabetes Care. 2004; 27(10): 2376-2381.
22. Monte FA, Melo PS, Alves A, Oliveira Junior JV, Alencar G, Soares FC. Evaluation of the southwick angle in two hundred hips of asymptomatic children and adolescents Rev Bras Ortop. 2020; 55(3): 360-366.

Address for correspondence

Ewa Puszczałowska-Lizis
University of Rzeszów
Collegium Medicum
Faculty of Health Sciences and Psychology
Rzeszów, Poland

Received: 18.10.2024
Accepted: 28.04.2025
Online publication: 11.03.2026