fortune Journals

Research Article

ARCHIVES OF PHYSIOTHERAPY AND REHABILITATION

Low Inter-rater Reliability of Active Hand Joint Range of Motion Measurements When Assessing Hand Function in a Clinical Setting

Margarida Vieira[1*], Alexander Kögel[2], Marie Stroetmann[2], Robert Wendlandt[3], Arndt-Peter Schulz[4,5]

Abstract

Background: Hand physiotherapists typically rely on goniometers for manual joint angle assessments, a process that is not only time-consuming but also susceptible to significant variability between therapists, which can lead to inconsistent evaluations, complicating the accurate tracking of a patient's progress. Inter-rater reliability tests are crucial in determining whether these standard assessment tools need to be improved or replaced, to achieve more accurate assessments and ultimately enhance the quality of patient care.

Methods: Participants in the study included five patients hospitalized in the hand department due to movement restrictions and five experienced hand physiotherapists. Following a standardized protocol, flexion and extension of all finger joints were measured, along with other thumb and wrist movements, using identical goniometers and a finger ruler to assess fingernail palm distance (FPD) and fingernail table distance (FNTD). Each joint measurement was performed once by each of the five therapists to evaluate inter-rater reliability.

Results: Results indicate varied inter-rater reliability, with 37.2% of measurements showing poor agreement (ICC < 0.5), 30.2% moderate agreement ($0.5 \leq$ ICC < 0.75), and 4.7% excellent agreement (ICC ≥ 0.9).

Conclusions: Findings underscore the need for improved measurement instruments in clinical settings, suggesting potential benefits from sensor gloves or other electrically based devices to enhance reliability and accuracy in ROM assessments.

Keywords: Hand; Physiotherapy; Goniometer; Treatment; Inter-rater reliability; Mobility; Range of motion; ICC

Introduction

Limitations in joint mobility can result from diverse factors such as medical conditions, fractures, inflammation, pain, accidents, among others. Hand injuries, despite a decline, still lead to the primary type of workplace injury. For instance, in 2021, according to the German Social Accident Insurance (DGUV), over 236,681 non-fatal work-related hand and wrist injuries occurred, leading to work absences longer than four days [1]. Such statistics underline the critical need for assessing the flexibility and movement of joints to evaluate functional mobility and musculoskeletal health, and it plays a significant role in choosing appropriate therapeutic treatments. The range of motion (ROM) is commonly categorized into Passive ROM (PROM), Active-assisted ROM (AAROM), and Active ROM (AROM), determining the extent

Affiliation:
[1]NOVA School of Science and Technology, Faculty of Biomedical engineering, Lisbon, Portugal

[2]Hand Rehabilitation Department, BG Klinikum Hamburg, Germany

[3]Biomechatronics Laboratory, Universitätsklinikum Schleswig-Holstein, Campus Lubeck, Germany

[4]Fraunhofer-Einrichtung für Individualisierte Medizintechnik IMTE, Lübeck, Germany

[5]Zentrum Klinische Forschung – ZKF, BG Klinikum Hamburg, Hamburg, Germany

***Corresponding author:**
Margarida Vieira, NOVA School of Science and Technology, Lisbon, Portugal

Citation: Margarida Vieira S, Alexander Kögel, Marie Stroetmann, Robert Wendlandt, Arndt-Peter Schulz. Low Inter-rater Reliability of Active Hand Joint Range of Motion Measurements When Assessing Hand Function in a Clinical Setting. Archives of Physiotherapy and Rehabilitation. 9 (2026): 14-19.

Received: December 24, 2025
Accepted: December 31, 2025
Published: February 06, 2026

Vieira M, et al., Arch Physiother Rehabil 2026
DOI:10.26502/fapr0030

of movement a joint can achieve [2,3]. Ensuring reliable ROM measurements is essential for effective therapy planning and evaluation. Still, it depends on factors such as clinician expertise, patient health, instrument accuracy and adherence to measurement protocols. Universal goniometers are widely used in clinical settings, particularly for small joints like the hand and wrist, due to their simplicity and cost-effectiveness [4]. Despite their widespread use as the clinical standard ROM assessment [5], universal goniometers are prone to variability, ranging from two to seven degrees in joint angle measurements [6] and have a level of measurement error of five degrees for hand joints [7]. Moreover, this measurement instrument is very time-consuming and requires careful attention to procedures to minimize errors. Additionally, measurements taken by different therapists on the same patient might vary significantly, highlighting the inconsistency in ROM assessments even under standardized conditions. There is no ground truth measurement possible for patient's hand movement as this would require potentially harmful measures such as radiography. Still, accurate measures are important when assessing the success of therapeutical measures or documenting a lasting disability. The Intraclass Correlation Coefficient (ICC) is a widely utilized statistical tool in medical research and clinical settings, serving as a reliability index for evaluating the consistency and agreement of measurements, whether taken by different raters (inter-rater reliability) or by the same instrument under varying conditions (intra-rater reliability). This reliability is crucial for ensuring the validity and reproducibility of medical research findings [8]. Numerous studies have examined inter-rater reliability (IRR) based on the ICC across various applications, but few focus on all hand joints. McGee et al. [9]. focused on the reliability of thumb active and passive flexion ROM. Reissner et al [10] compared goniometry and 3D motion analysis of finger and wrist ROM. Hancock et al [11] demonstrated the reliability of different knee goniometry methods performed by three raters, indicating which method had better reliability. Lewis et al. [12] focused on the IRR for the middle finger ROM of healthy patients using a Rolyan finger goniometer, measured by seven raters, and focusing only on flexion. Engstrand et al.´s study [7] closely resembles this one, examining the IRR of ROM of finger joints in people with Dupuytren's disease, focusing on extension and flexion of only three joints. These examples highlight the many applications of the ICC, which is the most used metric for calculating reliability in various types of measurements. Like many other studies, this one evaluates the discrepancy in measurements among therapists when assessing patients' hand mobility using a universal goniometer to capture AROM of all hand joints in static poses. Additionally, it involves participants with various hand movement limitations and examines all finger and wrist movements.

Methadology and Methods

Participants

Participants included five patients, four female and one male, with movement restrictions due to several causes such as distal radius fracture with ulnar styloid avulsion, Scapholunate Ligament (SL) ligament lesions, open fractures, partial SL and Scapholunate Triquetral Ligament (SLT) ligament ruptures, and persistent exercise-induced insufficiency after Triangular Fibrocartilage Complex (TFCC) lesions. Inclusion criteria comprised patients with hand movement impairment of the hand department. Exclusion criteria encompassed patients with AROM limited by pain, such as severe Complex Regional Pain Syndrome (CRPS), and those with burns injuries. Additionally, patients with multiple issues affecting range of motion in different joints were excluded from the study. Measurements were taken by five physiotherapists with a minimum of 10 years of experience in hand physiotherapy. Study center is the rehabilitation department of a dedicated trauma hospital acting as a tertial referral center. The therapists had not previously assessed or treated the study participants.

Materials

Uniformity in measurement tools was maintained throughout the study by consistently using specific instruments for each type of measurement: a small arm goniometer for the finger joints, a long arm goniometer for wrist movements, and a finger ruler for measuring distances from the fingernail to the palm and table. These tools, along with a brochure describing the warm-up movements and a form where the therapists write the measurement values (each table with five forms, one per therapist), are all presented in Figure 1. The units of measurement of the goniometer are in degrees with a resolution of two degrees increments.

Figure 1: Material Setup

Methodology

Initial warm-up

Before conducting measurements, an initial warm-up session was performed for all patients. The warm-up program consisted of five wrist and five finger exercises, conducted in three sets with 20 repetitions each, as detailed in the patient

Citation: Margarida Vieira S, Alexander Kögel, Marie Stroetmann, Robert Wendlandt, Arndt-Peter Schulz. Low Inter-rater Reliability of Active Hand Joint Range of Motion Measurements When Assessing Hand Function in a Clinical Setting. Archives of Physiotherapy and Rehabilitation. 9 (2026): 14-19.

brochure. These exercises were performed under the guidance of one of the five therapists, ensuring consistency and proper execution.

Setup

The study schedule was set from 8:00 am to 10:00 am. Prior to measurements, no therapies or exercises were administered to ensure consistency. Each participant received a confirmation sheet outlining the procedures. To maintain confidentiality and reduce bias, measurement results were not disclosed audibly in front of the patient, and neither patient nor therapist commented on the results. Moreover, to ensure that therapists did not have access to each other's results, each completed form was immediately placed in a yellow envelope designated for the specific participant. Participant identification was drafted as "P1" to "P5" to ensure anonymity. Similarly, therapists were identified as "T1" to "T5". A separate table was maintained to pair therapists with patients, ensuring blinding during data collection. An independent supervisor oversaw the organization of participants to maintain the integrity of the blinding procedure. Therapists were briefed on the measurement setup and procedural steps before commencing. Each measurement session was allocated 15 minutes per patient to ensure thoroughness and adherence to the study protocol. All patients were seated consistently in a designated location, with each having an individual table arranged in a circular formation. Therapists were positioned on the outside of this circle and rotated clockwise after completing measurements for each participant. Each therapist performed measurements only once per participant to ensure consistency. The data processing and calculations were performed by an independent researcher not involved in the measurement process or treatment of the participants.

Calculation of Inter-rater Reliability

Inter-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC), specifically the two-way random effects model, ICC (2.1) [13]. The chosen approach for calculating this coefficient involved the following specifications:

- An Anova Data Analysis was used in Excel: Two-factor Without Replication
- Model: Two-Way Random Effects-this model accounts for both raters and subjects as sources of random effects.
- Type of Relationship: Absolute Agreement-focuses on absolute differences between ratings, regardless of the order in which subjects are ranked.
- Unit: Single Rater-utilizes ratings from individual raters exclusively for measurement purposes.
- The ICC values were interpreted as follows: ICC less than 0.5 indicates poor reliability, ICC between 0.5 and 0.75 suggests moderate reliability, ICC between 0.75 and 0.9 indicates good reliability, and ICC greater than 0.9 indicates excellent reliability[13]. These criteria were used to evaluate the consistency of measurements across the different rates.

Measurements

For the measurements, an extensive set of assessments was performed, seen in Figure 2. For the fingers II-V, both extension and flexion were measured for each joint: Metacarpophalangeal (MCP), Proximal Interphalangeal (PIP) and Distal Interphalangeal (DIP), resulting in six measurements per finger. The thumb was evaluated for radial abduction (rad. ABD), opposition (Opp) and extension and flexion of both MCP and IP joints, totaling six measurements. Wrist measurements included dorsal extension, palmar flexion, radial abduction (Rabd) and ulnar abduction (Uabd), as well as supination (Sup) and pronation (Pro) of the hand. Additionally, distances from the fingernail to the palm and to the table were assessed for the fingers II to V, having a total of 44 measurements. However, the thumb opposition was excluded from the calculations because one participant´s data was missing, resulting in 43 measurements being used for calculations.

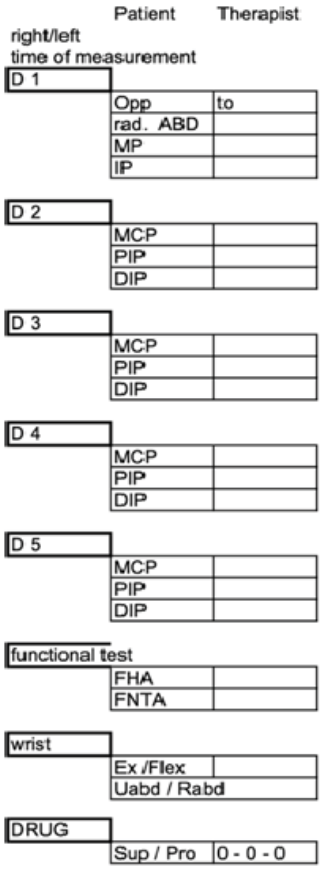

Figure 2: Example of each patient's form

Vieira M, et al., Arch Physiother Rehabil 2026
DOI:10.26502/fapr0030

Results

The table below presents the distribution of measurements categorized by their corresponding ICC values, while the ICC values for each measurement can be found in Table 1.

Table 1: Percentage of the measurements according to their ICC values.

ICC values	Measurements (%)
$ICC \geq 0.9$	4.7
$0.75 \leq ICC < 0.9$	27.9
$0.5 \leq ICC < 0.75$	30.2
$ICC < 0.5$	37.2

Figure 3 and Figure 4 show the differences in the ICC values between flexion and extension for each joint.

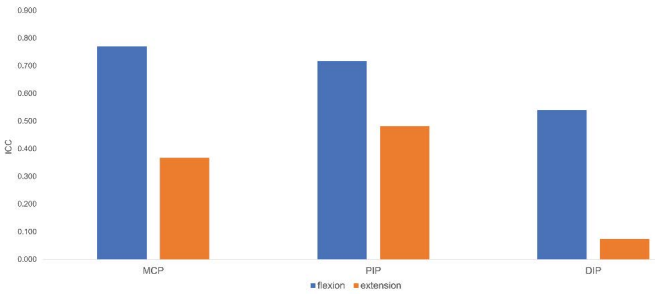

Figure 3: ICC values for the flexion and extension of the digits II-V joints

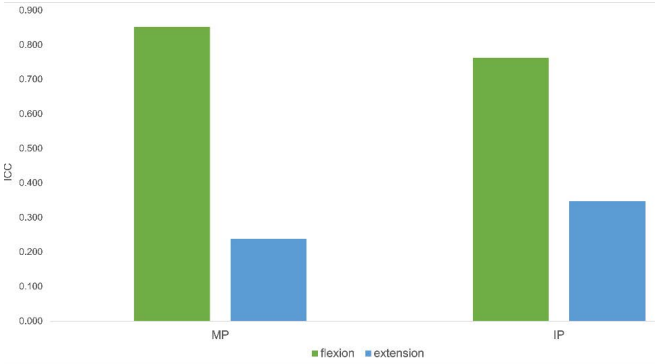

Figure 4: ICC values for the flexion and extension of the thumb joints

Discussion

The analysis of ICCs reveals varying agreement levels among raters using a goniometer. A notable finding is the substantial proportion (37.2%) of measurements with ICC values below 0.5, indicating poor agreement. Additionally, moderate agreement (30.2%) falls within the range of 0.5 to 0.75, suggesting ongoing variability in outcomes. While some measurements show excellent agreement (ICC ≥ 0.9), this proportion is small (4.7%). It is noteworthy that joint flexion consistently shows a higher ICC value compared to extension, likely due to the placement of the goniometer. Movements demonstrating higher reliability include MCP flexion of all fingers (ICC > 0.75), except for MCP5, and wrist supination, dorsal extension, palmar flexion, and ulnar abduction, all with ICC values > 0.8. Conversely, movements with lower reliability include extension of all finger joints (ICC < 0.5), except for PIP4 and PIP5. It is important to note that DIP flexion reliability for all fingers is the lowest compared to other joints' flexion, and DIP extension reliability for all fingers is also low (ICC < 0.2). This aligns with Lewis et al [12], who discussed the reliability of middle finger joint flexion in healthy adults using a Rolyan finger goniometer, showing that DIP joints are more difficult to measure due to the goniometer's shorter lever arms and the difficulty in manipulating them. However, the reliability of DIP in our study is even lower, likely due to the participants having movement limitations, unlike those in the study by Lewis et al. Finally, wrist movements with lower reliability include radial abduction (0.137) and pronation (0.256). These differences can be attributed to the more challenging and less consistent placement of the goniometer during extension, due to the positioning of the hand and the structures involved. McGee et al [9] attributed these observed differences to several factors, including the visualization of bony landmarks, the homogeneity of ROM in healthy joints, and concomitant joint impairments that can complicate the correct placement of the goniometer, particularly in participants with pathological joint constraints. While the observed variances among raters are high, these findings align with expectations considering goniometer limitations and the nature of the patients under assessment. Firstly, it's crucial to acknowledge that having a small sample size generally leads to less reliable estimates of ICC. Specifically, certain ICC results appeared unexpected due to the disproportionate influence of one whose assessments differed from the others. This disparity highlighted the sensitivity of ICC to small sample sizes. This aspect was mentioned by McGee et al [9], who noted that their study required 30 subjects to achieve sufficient statistical power. Therefore, a key recommendation for future studies is to increase the number of participants, with a minimum of 20 participants [14]. Moreover, our patient cohort comprises individuals with different hand and/or finger movement restrictions, leading to variability in their ROM over time. Despite the initial warm-up session, the patient's hand movement variability persisted. Consequently, these fluctuations may have contributed to the observed discrepancies in the ratings provided by different therapists. Engstrand et al [7] discussed this issue while investigating the inter-rater reliability of ROM in the finger joints of individuals with Dupuytren's Disease. They noted that repeating the same motion several times could influence the results due to factors such as learning, motivation, or fatigue, which can lead to a systematic change in the average results. Furthermore, the use of a goniometer, although a widely accepted tool for measuring ROM, may introduce imprecision into the

measurements. Reissner et al [10] found, by comparing the repeatability of ROM measurements of hand joints using a manual goniometer and a 3D motion capture system, that the reliability criterion (ICC > 0.7) was met for 94% of the joints with the 3D motion capture system but only for 65% with the goniometer. Also, averaged across all analyzed joints, the mean difference in degrees of variation was 10 degrees and 18 degrees for the 3D motion and goniometer, respectively, showing the incapability of the goniometer to detect smaller changes in joint mobility, which is crucial in clinical settings [6]. This study concluded that the goniometer lacks precision. Another study about knee goniometry conducted by Hancock et al [11] showed that the minimal significant differences for the long-arm goniometer and short-arm goniometer are 10 degrees and 14 degrees, respectively. This highlights, again, its limitations in detecting small changes in joint mobility.

Conclusion

This study highlights the varied inter-rater reliability of hand joint AROM measurements using a universal goniometer among patients with hand movement impairments. The findings reveal significant discrepancies, with a notable proportion of measurements showing poor agreement (37.2% with ICC < 0.5) and only a small fraction demonstrating excellent agreement (4.7% with ICC ≥ 0.9). The results emphasize the limitations of goniometers, particularly in detecting small changes in joint mobility and maintaining consistency between raters. For these reasons, different measurement instruments are needed to improve measurement consistency. With the advance of technology, sensor gloves and infra-red camera systems are already being researched to replace traditional methods [15,16], revolutionizing assessments. These instruments could outperform traditional methods by minimizing discrepancies among raters and mitigating subjective interpretations, resulting in reduced errors, improving therapeutic assessment and treatment planning.

Statements

Declaration of conflicting interest

The authors declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article

Ethical Approval

This study was approved by the ethic commission from the University of Lübeck (approval no. 2024-428).

Statement of Human and Animal Rights

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5). Informed consent was obtained from all patients to be included in the study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Acknowledgments

We would like to express our gratitude to the BG Klinikum Hamburg hand therapists, who participated in this study and contributed to its success: Laura Steinhauser, Benigna Giensch, Monika Schwiebert, and Birgit Maak. Their expertise and support played a crucial role in completing this research.

References

1. Gesetzliche Unfallversicherung Deutschland. Statistik – Arbeitsunfallgeschehen (2021).
2. Aly Yakout R, Khlosy H. Range of Motion Exercises Effect during and after Hydrotherapy on Patients Burned Hand Function and Pain Intensity: A Comparative Study. Egyptian Journal of Health Care 11 (2020): 670-687.
3. Kent, M. (2006). The Oxford Dictionary of Sports Science & Medicine (Third). Oxford University Press. https://doi.org/10.1093/acref/9780198568506.001.0001.
4. Fraeulin L, Holzgreve F, Brinkbäumer M, et al. Intra- and inter-rater reliability of joint range of motion tests using tape measure, digital inclinometer and inertial motion capturing. PLoS One 15 (2020): e0243646.
5. Hanks J, Myers B. Validity, Reliability, and Efficiency of a Standard Goniometer, Medical Inclinometer, and Builder's Inclinometer. Int J Sports Phys Ther 18 (2023): 989-996.
6. Reissner L, Fischer G, List R, et al. Minimal detectable difference of the finger and wrist range of motion: Comparison of goniometry and 3D motion analysis. J Orthop Surg Res 14 (2019).
7. Engstrand C, Krevers B, Kvist J. Interrater reliability in finger joint goniometer measurement in Dupuytren's disease. American Journal of Occupational Therapy 66 (2012): 98-103.
8. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol. 8 (2012): 23-34.
9. Mcgee C, Carlson K, Koethe A, et al. Inter-rater and inter-instrument reliability of goniometric thumb active and passive flexion range of motion measurements in healthy hands. Hand Ther 22 (2017): 175899831769075.

10. Reissner L, Fischer G, List R, et al. Minimal detectable difference of the finger and wrist range of motion: Comparison of goniometry and 3D motion analysis. J Orthop Surg Res 14 (2019): 173.

11. Hancock GE, Hepworth T, Wembridge K. Accuracy and reliability of knee goniometry methods. J Exp Orthop 5 (2018): 46.

12. Lewis E, Fors L, Tharion WJ. Interrater and intrarater reliability of finger goniometrie measurements. American Journal of Occupational Therapy 64 (2010): 555-561.

13. Meijer HAW, Graafland M, Obdeijn MC, et al. Validity and reliability of a wearable-controlled serious game and goniometer for telemonitoring of wrist fracture rehabilitation. European Journal of Trauma and Emergency Surgery 48 (2022): 1317-1325.

14. Koo T K, M Y Li. "Cracking the code: providing insight into the fundamentals of research and evidence-based practice a guideline of selecting and reporting intraclass correlation coefficients for reliability research." J Chiropr Med 15.2 (2016): 155-163.

15. Dicky O, Prima A, Ono Y, et al. Evaluation of Joint Range of Motion Measured by Vision Cameras (2019).

16. Lin BS, Lee IJ, Yang SY, et al. Design of an Inertial-Sensor-Based Data Glove for Hand Function Evaluation. Published online 18 (2018): 1545.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license 4.0

Citation: Margarida Vieira S, Alexander Kögel, Marie Stroetmann, Robert Wendlandt, Arndt-Peter Schulz. Low Inter-rater Reliability of Active Hand Joint Range of Motion Measurements When Assessing Hand Function in a Clinical Setting. Archives of Physiotherapy and Rehabilitation. 9 (2026): 14-19.