JSES International 7 (2023) 2528–2533

Contents lists available at ScienceDirect

JSES International

journal homepage: www.jsesinternational.org

# Remote visual estimation of shoulder range of motion has generally high interobserver reliability but limited accuracy

Simon Hwang, MD[a], Javier Ardebol, MD[a], Kassem Ghayyad, MD[a], Theresa Pak, DO[a], Joao A. Bonadiman, MD[b,c], Patrick J. Denard, MD[a,*], Mariano E. Menendez, MD[a], PacWest Shoulder Study Group[1]

[a]*Oregon Shoulder Institute, Medford, OR, USA*
[b]*Instituto de Ortopedia e Traumatologia, Hospital São Vicente de Paulo, Passo Fundo, RS, Brazil*
[c]*Instituto Brasil de Tecnologias da Saúde, Rio de Janeiro, RJ, Brazil*

ARTICLE INFO

*Keywords:*
Shoulder range of motion
Visual estimation
Telehealth
interobserver reliability
Intraclass correlation coefficient

*Level of evidence:* Basic Science Study; Validation of Outcome Assessment

**Background:** Surgeon visual estimation of shoulder range of motion (ROM) is commonplace in the outpatient office setting and routinely reported in clinical research, but the reliability and accuracy of this practice remain unclear. The purpose of this study is to establish the reliability and accuracy of remote visual estimation of shoulder ROM in healthy volunteers and symptomatic patients among a large group of shoulder surgeons. Our hypothesis is that remote visual estimation would be reliable and accurate compared with the digital goniometer method.

**Methods:** Fifty shoulder surgeon members of the PacWest Shoulder and Elbow Society independently determined the active shoulder forward flexion (FF), internal rotation at 90° abduction (IR90), external rotation at 90° abduction, external rotation at the side , and maximal spinal level reached with internal rotation ($IR_{spine}$) through visual estimation of video recordings taken from 10 healthy volunteers and 10 symptomatic patients. Variations in measurements were quantified using the interobserver reliability through calculation of the intraclass correlation coefficient. Accuracy was determined through comparison with digital goniometer measurements obtained with an on-screen protractor application using Bland–Altman mean differences and 95% limits of agreement.

**Results:** The interobserver reliability among examiners showed moderate to excellent correlation, with intraclass correlation coefficient ranging from 0.768 to 0.928 for the healthy volunteers and 0.739 to 0.878 for the symptomatic patients. Accuracy was limited, with upper limits of agreement exceeding the established minimal clinically important differences (MCIDs) for FF (20° vs. MCID of 14°) and IR90 (25° vs. 18°) in the healthy volunteers and for FF (33° vs. 16°), external rotation at 90° abduction (21° vs. 18°), and IR90 (31° vs. 20°) in the symptomatic patients.

**Conclusion:** Despite generally high intersurgeon reliability in the visual estimation of shoulder ROM, there was questionable accuracy when compared to digital goniometer measurements,with measurement errors often exceeding established MCID values. Given the potential implications for the clinical response to treatment and the significance of research findings, the adoption of validated instruments to measure ROM and the standardization of examination procedures should be considered.



Institutional review board approval was obtained prior the study. Southern Oregon IRB (Study number 22-007).

*Corresponding author: Patrick J. Denard, MD, Southern Oregon Orthopedics, 2780 E. Barnett Rd, Suite 200, Medford, OR 97530, USA.
*E-mail address:* pjdenard@gmail.com (P.J. Denard).

[1] PacWest Shoulder Study Group: Michael H. Amini, MD; Anup Shah, MD; Andy Hartman, MD; Anthony Wei, MD; Brian Lee, MD; Benjamin W. Sears, MD; Christopher Sheu, MD; Christopher R Lehman, MD; Daniel J. Solomon, MD; David Weinstein, MD; Dawson S. Brown, MD; David Savin, MD; Dave Huberty, MD; Dan Guttmann, MD; Dave Shukla, MD; Edward Choung, MD; Evan Lederman, MD; Patrick J Denard, MD; Mariano E Menendez, MD; Elizabeth Watson, MD; Erica M. Burns, MD; Geoffrey Abrams, MD; Gregory D. Gramstad, MD; Samuel Harmsen, MD; Hafiz Kassam, MD; Jae You, MD; Jason Kurian, MD; Jason R. Saleh, MD; Jonah Davies, MD; Joseph W. Galvin, DO; Raffy Mirzayan, MD; Liang Zhou, MD; Matthew Budge, MD; Michael A. Stone, MD; Matthew Nugent, MD; Matthew Pifer, MD; Ron Navarro, MD; Niloofar Dehghan, MD; Nathan Orvets, MD; Reza Omid, MD; Robert Orfaly, MD; Peter Chalmers, MD; Peters T. Otlans, MD, MPH; Rachel M. Frank, MD; Rudolf G. Hoellrich, MD; Scott Humphrey, MD; Scott Crow, MD; Shanjean Lee, MD; Richard Lee Stowell, MD; Ekaterina Urch, MD; William Paterson, MD; John Costouros, MD; Angela Passanise, DO; Anita G Rao, MD; Armodios Hatzidakis, MD; J Brad Butler, MD; Cay Mierisch, MD; Sara Jurek, MD; Hilary Malcarney, MD; Jason Hsu, MD; Jeffrey Tuman, MD; Jonathan Cheah, MD; John Matthews, MD; Kyong Min, MD; Lauren Hansen, MD; Julia Lee, MD; Raymond Klug, MD; Rafael Buerba, MD, MHS; Timothy Crall, MD; Todd Kim, MD; Yassamin Hazrati, MD.

https://doi.org/10.1016/j.jseint.2023.07.002

Measurements of range of motion (ROM) are widely utilized for the functional evaluation of the shoulder joint and play a vital role in many aspects of clinical decision-making and research. Reliability and accuracy of these measurements are important to ensure that a change seen in the patient's status following operative or nonoperative intervention is genuine rather than an error in measuring methods. Goniometry and visual estimation are the two most frequently employed methods for evaluating ROM.[7,12,27] Clinically, visual estimation is more commonly utilized due to its efficiency and practicality, despite lower reliability and accuracy when compared to goniometer-based measurements.[12,27,33] While visual estimation may be adequate for a single practitioner to monitor a patient's longitudinal progress, studies have shown low interobserver reliability, particularly in cases involving shoulder pathology or pain.[23,31,33] This is relevant in the current health care environment, as patients are often assessed by residents, clinical fellows, physician assistants, or primary care physicians prior to their clinical evaluation by an orthopedic surgeon.

Accelerated by the COVID-19 pandemic, the expanding role of telehealth has benefited the orthopedic specialty by increasing access to care while providing safe and cost-effective care.[5,11,6] However, concerns have been raised regarding the feasibility of performing a comprehensive orthopedic physical examination during a virtual clinic encounter.[19,20] Inherent limitations of the virtual examination may impact the reliability and accuracy of ROM measurements obtained over a telehealth video. However, a variety of methods have been validated for remote assessment of ROM, including still photography,[9,21] smartphone clinometers,[16,22,37] motion-based software development kits,[25] and on-screen protractor applications.[26,28] While these validated tools demonstrate excellent inter- and intraobserver reliability, few studies have examined the reliability of visual estimation in shoulder ROM assessment, especially in the virtual setting.

The purpose of this study is to establish the reliability and accuracy of remote visual estimation of shoulder ROM in healthy volunteers and symptomatic patients among a large group of shoulder surgeons. Our hypothesis is that remote visual estimation would be reliable and accurate compared with the digital goniometer method.

## Materials and methods

### *Subjects*

A convenience sample of ten healthy volunteers was recruited for participation in the initial study over the course of 1 week. All volunteers were either health care professionals or research fellows at the authors' institution, aged between 23 and 46 years (mean, 35.2 years), and able to perform full painless ROM in both shoulders. There were 8 men and 2 women. The right shoulder was examined in 6 subjects, and the left shoulder was examined in 4.

Following completion of the initial study, ten symptomatic patients (6 men and 4 women) were recruited for evaluation. All patients in this arm of the study were seen in the office for either glenohumeral arthritis or rotator cuff pathology but were examined prior to any surgical intervention of the pathologic shoulder. The right shoulder was examined in 6 patients, and the left shoulder was examined in 4. Following consent, basic instructions on how to perform the shoulder ROM maneuvers were provided to each patient.

### *Surgeon evaluation*

To mimic the setting of a telehealth visit, all subjects were recorded with a camera phone (iPhone SE; Apple, Cupertino, CA, USA) as they performed the required shoulder maneuvers. The ROM videos of all subjects were distributed via an email survey developed using Google Forms to 180 members of the PacWest Shoulder and Elbow Society. Formed in 2021, this society aims to facilitate the exchange of ideas and research collaboration and includes shoulder surgeons from the following states: Alaska, Washington, Idaho, Oregon, Nevada, California, Arizona, Hawaii, Colorado, New Mexico, Utah, Montana, and Wyoming. The ROM measurements from both sets of subjects (healthy and symptomatic) were evaluated through remote visual estimation by members of this society.

### *Digital goniometer*

Three independent reviewers (S.H., J.A, and K.G.) obtained digital goniometer measurements on the same set of videos using an on-screen protractor application (Fig. 1). This validated method has been shown to reliably and accurately assess video shoulder ROM.[28] The average measurement from the three reviewers was calculated for each maneuver in each subject, and these values were then compared with the visual estimates provided by the surgeons.

### *Shoulder ROM measurement of healthy volunteers*

Randomized ROM targets for each shoulder maneuver were assigned to each healthy subject. These targets were established within the established normal ROM values for shoulder forward flexion ([FF]; 0°-180°), external rotation at 90° abduction ([ER90]; 0°-90°), internal rotation at 90° abduction ([IR90]; 0°-70°), external rotation at 0° abduction ([ER0]; 0°-70°), and maximal spinal level reached with internal rotation ([$IR_{spine}$]; hip-T1).[32] Each subject was instructed to perform this consecutive series of maneuvers (FF, ER90, IR90, ER0, and $IR_{spine}$), pausing briefly at the assigned ROM target before returning to the neutral position. The intent of this procedure was to simulate ROM deficits typically observed on examination of a pathologic shoulder. Video recordings of each subject were taken with the camera lens positioned parallel to the sagittal plane at approximately chest level. Blinded videos were then distributed to the PacWest shoulder study group, and remote visual estimation measurements were collected. Three independent reviewers (S.H., J.A, and K.G.) evaluated the videos with a digital goniometer "protractor" application that is available as an extension on the Chrome browser. All examiners were blinded to each other's measurement data, and videos were removed from the survey following 2 weeks of response collection.

### *Shoulder ROM measurement of symptomatic patients*

Ten patients with glenohumeral arthritis or rotator cuff pathology and expected deficits in shoulder ROM were recruited from the clinic. The patients were instructed to move the upper extremity to the end of active FF, ER90, IR90, ER0, and $IR_{spine}$ within the comfort levels of their pathologic shoulder. The ROM maneuvers were recorded, and the blinded patient videos were distributed to the same group of surgeon examiners for remote visual estimation. Similar to procedures used in healthy volunteers, videos of the symptomatic patients were evaluated with the on-screen protractor.

### *Statistical analysis*

Our study defined "reliability" as the extent to which a measure generates consistent outcomes when administered by multiple

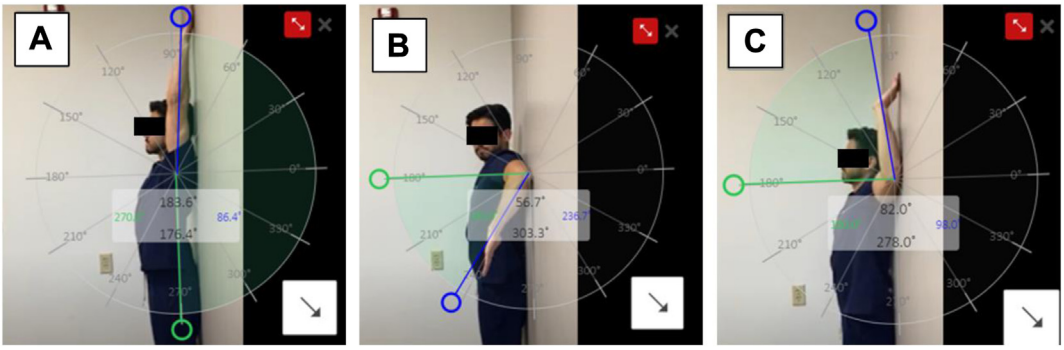


**Figure 1** Digital goniometer measurements from a validated, on-screen protractor application.[28] (A) forward flexion. (B) internal rotation at 90° abduction. (C) external rotation at 90° abduction.

raters. High reliability implies that if measurements were to be repeated by different raters, it would yield similar results. On the other hand, "accuracy" refers to the degree to which the measurements reflect the true values or characteristics of the population being studied. While reliability is expressed as a correlation coefficient that lacks units, accuracy is measured in the same units as the original measurement.

In order to determine the reliability of remote visual estimation of shoulder ROM among orthopedic surgeons in healthy volunteers and symptomatic patients, the interobserver reliability was calculated with the intraclass correlation coefficient (ICC) using a 2-way random-effect model with a single measure and absolute agreement among measurements for each maneuver. The ICC for each maneuver was expressed with a 95% confidence interval. The ICC values were used to determine the interobserver reliability (1) between the 50 surgeon examiners for the remote visual estimation of shoulder ROM in healthy volunteers and symptomatic patients, (2) averaged from the individual ICC values between the visual estimates and on-screen protractor measurements, and (3) among the 3 medical doctor examiners for the on-screen protractor. Interpretation of the ICC is based on the definition provided by Koo and Li: 0.00-0.50, poor correlation; 0.50-0.75 moderate correlation; 0.75-0.90, good correlation; and 0.90-1.00, excellent correlation.[18] Sample size estimation was conducted at a power = 80% and an alpha = 0.05. In order to detect an ICC greater than 0.2 with 50 observations per subject, a minimum of 4 subjects were needed.[4]

To determine the accuracy of remote visual estimation, the data was analyzed by Bland–Altman analysis with 95% limits of agreement (LOA). This method is based on the mean and standard deviation of the difference between two measurements of the same subject.[2] The differences between the surgeon estimates and the on-screen protractor measurements were calculated for each shoulder maneuver and for each examiner. These differences were then averaged to calculate mean differences along with 95% LOAs expressed in degrees. The 95% LOA, which estimates the size of the possible sampling error, was calculated as 1.96× SD, in which SD is the standard deviation of the mean differences for each maneuver and each examiner. The 95% LOA values were compared with the established minimum clinically important differences (MCIDs) for shoulder ROM to determine the acceptability of the visual estimates.[23]

The standard error of measurement (SEM) was calculated for each maneuver as an additional measure of absolute reliability. The formula used to calculate SEM as described by Atkinson and Nevill[1] is SEM = SD $\sqrt{}$(1-ICC), where *SD* is the standard deviation and *ICC* is the ICC of each shoulder maneuver. The result of this formula is a SEM value expressed in degrees.

**Table I**
Interobserver reliability of visual estimation in healthy subjects and symptomatic patients.

| Measurement | ICC | 95% CI | SEM |
|---|---|---|---|
| Healthy | | | |
| FF | 0.928 | 0.857-0.977 | 11° |
| ER90 | 0.900 | 0.808-0.968 | 8° |
| IR90 | 0.768 | 0.604-0.918 | 8° |
| ER0 | 0.850 | 0.723-0.950 | 8° |
| $IR_{spine}$ | 0.889 | 0.787-0.964 | 2 (spinal levels) |
| Symptomatic | | | |
| FF | 0.878 | 0.769-0.960 | 12° |
| ER90 | 0.866 | 0.749-0.956 | 9° |
| IR90 | 0.739 | 0.565-0.905 | 12° |
| ER0 | 0.818 | 0.672-0.938 | 9° |
| $IR_{spine}$ | 0.795 | 0.641-0.929 | 2 (spinal levels) |

*ICC*, intraclass coefficient; *CI*, confidence interval; *SEM*, standard error of measurement; *FF*, forward flexion; *ER90*, external rotation at 90° abduction; *IR90*, internal rotation at 90° abduction; *ER0*, external rotation at 0° abduction; *IRspine*, maximal spinal level reached on internal rotation.

## Results

Fifty shoulder surgeons completed the surveys. Of these respondents, the majority indicated completion of either a shoulder and elbow fellowship (25; 50%) or a sports medicine fellowship (17; 34%). Forty (80%) surgeons indicated being in practice for ≥6 years. The mean number of approximate shoulder surgeries performed per year by respondents was 295 (range: 50-850).

### *Interobserver reliability in visual estimation measurements of healthy volunteers and symptomatic patients*

Overall, ICC values for all maneuvers (FF, ER90, IR90, ER0, and $IR_{spine}$) were greater among the healthy volunteers when compared to the symptomatic patients (Table I). In the healthy volunteers, ICC values were excellent for FF (0.928) and ER90 (0.900). A good correlation was seen for IR90 (0.768), ER0 (0.850), and $IR_{spine}$ (0.889). Among symptomatic patients, the ICC values were good for FF (0.878), ER90 (0.866), ER0 abduction (0.818), and $IR_{spine}$ (0.795). Only a moderate correlation was seen for IR90 (0.739).

SEM calculations for each maneuver were used to determine the absolute validity (accuracy). SEM values were greater among symptomatic patients compared to the healthy volunteers, ranging between 9°-12° in symptomatic patients and 8°-11° in healthy volunteers. In general, FF and IR90 exhibited the highest (worst) SEM. The calculated SEM for $IR_{spine}$ was 2 spinal levels in both cohorts (healthy and symptomatic).

*Accuracy of visual estimation measurements compared with digital goniometer in healthy volunteers and symptomatic patients*

For healthy volunteers, mean ICCs ranged from moderate to excellent (Table II). Table III demonstrates the interobserver reliability of the digital goniometer measurements among the three graduate medical doctors, which shows near perfect ICC values for FF (0.995), ER90 (0.984), and IR90 (0.951). Comparison of the remote visual estimation measurements with the digital goniometer measurements demonstrated excellent correlation for FF (0.958) and ER90 (0.923), and only moderate correlation for IR90 (0.703). Bland–Altman analysis showed minor mean differences for healthy volunteers in FF (2°) and ER90 (1°) and a larger difference for IR90 (9°). The upper 95% LOA was above the established MCID for FF (20° vs. 14°) and IR90 (25° vs. 18°) and below the MCID for ER90 (17° vs. 18°).

For symptomatic patients, the mean ICCs for FF (0.881 vs. 0.958), ER90 (0.878 vs. 0.923), and IR90 (0.700 vs. 0.703) were lower when compared to the healthy volunteers. Interobserver reliability of the digital goniometer once again demonstrated near-perfect ICC values for FF (0.961), ER90 (0.985), and IR90 (0.969). Comparison of the visual ROM estimates with the corresponding digital goniometer measurements demonstrated good correlation for FF (0.881) and ER90 abduction (0.878) and only moderate correlation for IR90 (0.700). Bland-Altman showed minor mean difference for healthy volunteers in ER90 (1°) and larger differences for FF (8°) and IR90 (8°). The upper 95% LOA was above the established MCID for FF (33° vs. 16°), ER90 (21° vs. 18°), and IR90 (31° vs. 20°).

In both groups, visual estimates for FF and ER90 exhibited greater accuracy when the arc of shoulder movement was approximately 90° (±10°). Surgeon estimates of shoulder ROMs in this range were within a 5° margin of the digital goniometer measurements in 87% (262/300) of cases.

## Discussion

The results of this study demonstrate moderate to excellent interobserver reliability of the visual estimation of shoulder ROM among a large cohort of surgeon examiners. The interobserver reliability of all shoulder maneuvers, as measured by the ICC, averaged 0.867 (range: 0.768-0.928) among the healthy volunteers. In the symptomatic group, the average ICC across all maneuvers was 0.819 (range: 0.739-0.878). IR90 exhibited the lowest ICCs with the widest confidence intervals, while FF and ER90 demonstrated the greatest interobserver reliability. The visual estimates demonstrated limited accuracy when compared to digital goniometer measurements. The average SEM across all maneuvers was 9° (range, 8°-11°) in the healthy group and 11° (range, 9°-12°) in the symptomatic cohort. Mean differences averaged 4° (range, 1°-9°), with 95% LOAs between 16° and 18° among the healthy group. In the symptomatic cohort, mean differences averaged 6° (range, 1°-8°) with 95% LOAs between 20° and 25°.

Reliable and accurate measurement of ROM is important for the objective assessment of shoulder function and affects the diagnosis, treatment, and scientific investigation of shoulder pathologies. Accelerated by the COVID-19 pandemic, the adoption of telemedicine in health care has impacted the feasibility of the traditional in-office physical examination.[11] Many clinical tools have been developed for the remote assessment of shoulder ROM, including limb-tracking software, photography-based goniometry, and smartphone clinometers.[3,9,37] However, assessment of shoulder ROM through visual estimation remains widely utilized due to its convenience and practicality. Concerns over the reliability and accuracy of remote visual estimation have been posed due to challenges inherent to the telemedicine platform.[29,35] Accuracy in documentation has major implications for both clinical care and research. In this study, we aim to define the reliability and accuracy of shoulder ROM measurements obtained from video-based, visual estimation by a large group of shoulder surgeons in both healthy volunteers and symptomatic patients.

Several studies have shown a wide range of interobserver reliability for the visual estimation of shoulder ROM.[13,14,31,33,35,37] Werner et al reported ICCs of 0.48 to 0.90 in a study of 24 healthy volunteers and 5 examiners with varying levels of expertise.[37] Several additional studies have reported ICCs for in-office visual estimation of shoulder ROM including Hayes et al (ICC range, 0.57-0.70), Hall et al (ICC range, 0.57-0.96), and Croft et al (ICC range, 0.43-0.95).[8,13,14] One prior study examined the interobserver reliability of telemedicine shoulder examinations utilizing the Kuder-Richardson formula 20 (KR-20), interpreted as: <0.5, unacceptable reliability; 0.5-0.6, poor reliability; 0.6-0.7, questionable reliability; 0.7-0.8, acceptable reliability; 0.8-0.9, good reliability; ≥0.9, excellent reliability.[35] The authors reported unacceptable to questionable reliability for FF (KR-20, 0.369), external rotation (KR-20, 0.518), and internal rotation (KR-20, 0.662).[35] The interobserver reliability in our study ranged from moderate to excellent, which was higher than what has been previously reported.[14,35,37] A key difference in our study is the incorporation of similarly-skilled, expert examiners rather than examiners with varying skill levels. As poor reliability has been demonstrated among nonphysician examiners, our cohort comprised of only orthopedic shoulder surgeons may have contributed to the greater interobserver reliability seen in our study.[35]

Interestingly, IR90 had the lowest interobserver reliability of all shoulder maneuvers in both healthy volunteers (ICC, 0.768; 95% CI [0.604-0.918]) and symptomatic patients (ICC, 0.739; 95% CI [0.565-0.905]). Prior studies have similarly shown low ICCs for IR90 in comparison to other planes of shoulder motion.[15,23,24,37] Werner et al demonstrated that among symptomatic patients, IR90 ICCs were 0.48 (95% CI, 0.22-0.63) for visual estimation and 0.56 (95% CI, 0.28-0.72) for standard goniometer measurements, which were the lowest among all tested shoulder movements.[37] Comparatively low IR90 ICCs obtained from standard goniometer measurements have been reported by Muir et al (ICC range, 0.39-0.62) and Mullaney et al (ICC range, 0.62-0.87).[23,24] As ICC values below 0.75 indicate poor to moderate reliability, findings from our study and prior studies indicate that both visual estimation and goniometer measurements of IR90 may be inconsistent if performed by different examiners.[18]

Our study also assessed the accuracy of visual estimation compared to a digital goniometer measurement, as a goniometer is often considered the gold standard in ROM measurement. Use of the digital goniometer by 3 physicians demonstrated near-perfect interobserver reliability (ICC range, 0.951-0.995; average ICC, 0.974). Comparison of the visual estimates with the digital goniometer measurements demonstrated a mean ICC of 0.861 (range, 0.703-0.958) in healthy volunteers and 0.820 (range, 0.700-0.881) in symptomatic patients. While these findings indicate good interobserver reliability, the values should be carefully interpreted as ICCs are influenced by between-subject variability.[36] We therefore calculated the SEM and Bland–Altman mean differences with 95% LOAs as measures of absolute reliability.[2,37] Overall, the absolute reliability (accuracy) of video-based, visual estimation of shoulder ROM was found to be poor. The average SEM among the healthy volunteers and symptomatic patients was 9° (range, 8°-11°) and 11° (range, 9°-12°), respectively. Mean differences between the visual estimates and digital goniometer measurements ranged from 1° to 9°, with 95% LOAs between 16° and 18° among the healthy volunteers. In the symptomatic patients, mean differences ranged from 1° to 8°, with 95% LOAs between 20° and 25°.

**Table II**
Comparison of remote visual estimation measurements and validated digital goniometer (on-screen protractor) measurements in healthy subjects and symptomatic patients: ICC, Bland-Altman mean differences, 95% Limits of agreement.

| Measurement | ICC | 95% CI | Mean difference (s.d.), ° | LOA (s.d.), ° | MCID,[23] ° |
|---|---|---|---|---|---|
| Healthy | | | | | |
| FF | 0.958 | 0.936-0.981 | 2 (6) | 18 (11) | 14 |
| ER90 | 0.923 | 0.919-0.937 | 1 (4) | 16 (5) | 18 |
| IR90 | 0.703 | 0.679-0.736 | 9 (5) | 16 (6) | 18 |
| Symptomatic | | | | | |
| FF | 0.881 | 0.854-0.908 | 8 (7) | 25 (12) | 16 |
| ER90 | 0.878 | 0.851-0.906 | 1 (6) | 20 (7) | 18 |
| IR90 | 0.700 | 0.661-0.739 | 8 (8) | 23 (8) | 20 |

*ICC*, intraclass coefficient; *CI*, confidence interval; *s.d.*, standard deviation; *LOA*, limits of agreement; *MCID*, minimal clinically important difference; *FF*, forward flexion; *ER90*, external rotation at 90° abduction; *IR90*, internal rotation at 90° abduction; $IR_{spine}$, maximal spinal level reached on internal rotation.

**Table III**
Interobserver reliability of on-screen protractor measurements in healthy subjects and symptomatic patients.

| Measurement | ICC | 95% CI |
|---|---|---|
| Healthy | | |
| FF | 0.995 | 0.986-0.999 |
| ER90 | 0.984 | 0.955-0.996 |
| IR90 | 0.951 | 0.869-0.986 |
| Symptomatic | | |
| FF | 0.961 | 0.889-0.989 |
| ER90 | 0.985 | 0.951-0.996 |
| IR90 | 0.969 | 0.912-0.991 |

*ICC*, intraclass coefficient; *CI*, confidence interval; *FF*, forward flexion; *ER90*, external rotation at 90° abduction; *IR90*, internal rotation at 90° abduction.

From a clinical perspective, this cannot be overlooked; as the upper ranges of disagreement of 33°, 21°, and 31° for FF, ER90, and IR90, respectively, in the symptomatic cohort exceeded the established MCID values for each maneuver.[23] This raises concerns regarding the potential impact on the clinical management of shoulder pathologies and the interpretation of research findings. Potential factors contributing to increased variation in visual estimates of shoulder ROM over the telemedicine setting include: non-standardized positioning of the camera and patient, baggy or concealing clothing, and inability to stabilize the joint during each maneuver. Validated tools such as smartphone applications or digital goniometers have demonstrated excellent concurrent validity with the in-office goniometer.[28,30,37] Additionally, the use of standardized procedures during the virtual examination has shown to increase the reliability and accuracy of ROM evaluation.[10] Incorporation of these tools and techniques for the remote assessment of shoulder ROM can limit the variation in measurements seen in the virtual setting.

To our knowledge, this is the first study to examine the reliability and accuracy of visual estimation among a large cohort of expert clinicians. Our study is, however, not without limitations. The positioning of the camera and patient was not standardized. While this may have contributed to increased variation in measurement, our intent was to mimic the telehealth setting, where camera positioning is typically variable. Additionally, the setting that the surgeons evaluated the videos was not standardized. Our findings could have been influenced by the variation in type of device used and the timepoint at which the videos were evaluated by the surgeons. The average age of the healthy volunteers (mean age = 35.2 years) differed from that of the symptomatic patients (mean age = 57.9 years). Differences in age and symptomatic status of the shoulder may have contributed to the increased variation seen in the symptomatic cohort, a trend which is has been previously demonstrated in literature.[17,34] Finally, intraobserver reliability was not assessed in the current study. However, previous studies have already demonstrated greater consistency with intraobserver measurements compared to interobserver measurements.[13,23,24,31] The intention of our investigation was simply to examine the between-surgeon variation.

## Conclusion

Remote visual estimation of shoulder ROM demonstrated moderate to excellent agreement between a large group of shoulder surgeons. However, questionable accuracy was demonstrated when compared to the digital goniometer, with measurement error exceeding established MCID values for FF, ER90° abduction, and internal rotation at 90° abduction. Given the wide availability and low cost of validated methods of remote shoulder ROM assessment, incorporation of these tools into the virtual examination workflow can limit the variation inherent to the telemedicine setting.

## Disclaimers:

Funding: No funding was disclosed by the authors.
Conflicts of interest:
Dr. Denard is a consultant for Arthrex, Inc.
Dr. Menendez is a consultant for Arthrex, Inc.
Dr. Andy Hartman is a consultant for Arthrex, Inc.
Dr. Daniel J. Solomon is a consultant for Arthrex, Inc.
Dr. Ekaterina Urch is a consultant for Arthrex, Inc.
Dr. Evan Lederman is a consultant, receives royalties and research support from Arthrex, Inc.
Dr. Geoffrey Abrams is a consultant for Arthrex, Inc. and Bioventus Inc. and has stock options for Sparta Biomedical.
Dr. Hafiz Kassam is a consultant for Zimmer-Biomet, Smith and Nephew, Arthrex Inc. and Catalyst Orthoscience.
Dr. J Brad Butler is a consultant for DePuy Mitek.
Dr. Jason Kurian is a paid consultant for DePuy/Synthes/Mitek.
Dr. John Costouros is a consultant for Stryker, Inc.
Dr. Matthew Budge is a consultant for Catalyst Orthoscience, Inc.
Dr. Matthew Pifer is a consultant for Arthrex, Inc.
Dr. Rachel M. Frank is a consultant for Arthrex, JRF, AlloSource, BodyCAD, on the editorial board for Elsevier, and receives research support for Arthrex.
Dr. Raffy Mirzayan is a consultant for Arthrex, Inc.
Dr. Raymond Klug is a consultant for Exactech, Inc.
Dr. Samuel Harmsen is a consultant for Arthrex, Inc. Shoulder Innovations, Inc., Embody, Inc., Enovis, Inc., Zimmer US Inc., and Genesis Software Innovations.
Dr. Scott Humphrey receives royalties from DJO Surgical, is on the speakers bureau for DJO Surgical, is a paid consultant for DJO Surgical and United Orthopedics, and is a board member for the Idaho Shoulder and Elbow Research Foundation (ISERF).
Dr. Timothy Crall is a consultant for Arthrex, Inc.

Dr. William Paterson is a consultant for FX Shoulder Solutions. The other authors, their immediate families, and any research foundation with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article.

## References


1. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 1998;26:217-38.
2. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135-60.
3. Blonna D, Zarkadas PC, Fitzsimmons JS, O'Driscoll SW. Validation of a photography-based goniometry method for measuring joint range of motion. J Shoulder Elbow Surg 2012;21:29-35. https://doi.org/10.1016/j.jse.2011.06.018.
4. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Arch Orofac Sci 2017;12:1-11.
5. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-effectiveness of telemedicine in remote orthopedic consultations: Randomized controlled trial. J Med Internet Res 2019;21:e11330. https://doi.org/10.2196/11330.
6. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Patient reported outcomes with remote orthopaedic consultations by telemedicine: a randomised controlled trial. J Telemed Telecare 2019;25:451-9. https://doi.org/10.1177/1357633x18783921.
7. Cools AM, De Wilde L, Van Tongel A, Ceyssens C, Ryckewaert R, Cambier DC. Measuring shoulder external and internal rotation strength and range of motion: comprehensive intra-rater and inter-rater reliability study of several testing protocols. J Shoulder Elbow Surg 2014;23:1454-61. https://doi.org/10.1016/j.jse.2014.01.006.
8. Croft P, Pope D, Boswell R, Rigby A, Silman A. Observer variability in measuring elevation and external rotation of the shoulder. Primary care rheumatology society shoulder study group. Br J Rheumatol 1994;33:942-6.
9. Elrahim RMA, Embaby EA, Ali MF, Kamel RM. Inter-rater and intra-rater reliability of Kinovea software for measurement of shoulder range of motion. Bull Fac Phys Ther 2016;21:80-7. https://doi.org/10.4103/1110-6611.196778.
10. Fabrés Martín C, Ventura Parellada C, Herrero Antón de Vez H, Ordoñez Urgiles CE, Alonso-Rodríguez Piedra J, Mora Guix JM. Telemedicine approach for patient follow-up after total knee and reverse total shoulder arthroplasty: a pilot study. Int J Comput Assist Radiol Surg 2023;18:595-602. https://doi.org/10.1007/s11548-022-02784-z.
11. Foni NO, Costa LAV, Velloso LMR, Pedrotti CHS. Telemedicine: is it a tool for orthopedics? Curr Rev Musculoskelet Med 2020;13:797-801. https://doi.org/10.1007/s12178-020-09680-6.
12. Gajdosik RL, Bohannon RW. Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. Phys Ther 1987;67:1867-72.
13. Hall JM, Azar FM, Miller RH 3rd, Smith R, Throckmorton TW. Accuracy and reliability testing of two methods to measure internal rotation of the glenohumeral joint. J Shoulder Elbow Surg 2014;23:1296-300. https://doi.org/10.1016/j.jse.2013.12.015.
14. Hayes K, Walton JR, Szomor ZR, Murrell GA. Reliability of five methods for assessing shoulder range of motion. Aust J Physiother 2001;47:289-94.
15. Hoving JL, Buchbinder R, Green S, Forbes A, Bellamy N, Brand C, et al. How reliably do rheumatologists measure shoulder movement? Ann Rheum Dis 2002;61:612-6. https://doi.org/10.1136/ard.61.7.612.
16. Keijsers R, Zwerus EL, van Lith DRM, Koenraadt KLM, Goossens P, The B, et al. Validity and reliability of elbow range of motion measurements using digital Photographs, Movies, and a goniometry smartphone application. J Sports Med 2018;2018:7906875. https://doi.org/10.1155/2018/7906875.
17. Kolber MJ, Hanney WJ. The reliability and concurrent validity of shoulder mobility measurements using a digital inclinometer and goniometer: a technical report. Int J Sports Phys Ther 2012;7:306-13.
18. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155-63. https://doi.org/10.1016/j.jcm.2016.02.012.
19. Kumar S, Kumar A, Kumar M, Kumar A, Arora R, Sehrawat R. Feasibility of telemedicine in maintaining follow-up of orthopaedic patients and their satisfaction: a preliminary study. J Clin Orthop Trauma 2020;11:S704-10. https://doi.org/10.1016/j.jcot.2020.07.026.
20. McLiesh P. Telehealth in contemporary orthopaedic nursing. Int J Orthop Trauma Nurs 2019;33:1-3. https://doi.org/10.1016/j.ijotn.2019.02.003.
21. Meislin MA, Wagner ER, Shin AY. A comparison of elbow range of motion measurements: smartphone-based digital photography versus goniometric measurements. J Hand Surg Am 2016;41:510-515.e1. https://doi.org/10.1016/j.jhsa.2016.01.006.
22. Mejia-Hernandez K, Chang A, Eardley-Harris N, Jaarsma R, Gill TK, McLean JM. Smartphone applications for the evaluation of pathologic shoulder range of motion and shoulder scores-a comparative study. JSES Open Access 2018;2:109-14. https://doi.org/10.1016/j.jses.2017.10.001.
23. Muir SW, Corea CL, Beaupre L. Evaluating change in clinical status: reliability and measures of agreement for the assessment of glenohumeral range of motion. N Am J Sports Phys Ther 2010;5:98-110.
24. Mullaney MJ, McHugh MP, Johnson CP, Tyler TF. Reliability of shoulder range of motion comparing a goniometer to a digital level. Physiother Theory Pract 2010;26:327-33. https://doi.org/10.3109/09593980903094230.
25. Ramkumar PN, Haeberle HS, Navarro SM, Sultan AA, Mont MA, Ricchetti ET, et al. Mobile technology and telemedicine for shoulder range of motion: validation of a motion-based machine-learning software development kit. J Shoulder Elbow Surg 2018;27:1198-204. https://doi.org/10.1016/j.jse.2018.01.013.
26. Russell TG, Jull GA, Wootton R. Can the Internet be used as a medium to evaluate knee angle? Man Ther 2003;8:242-6. https://doi.org/10.1016/s1356-689x(03)00016-x.
27. Russo RR, Burn MB, Ismaily SK, Gerrie BJ, Han S, Alexander J, et al. Is digital photography an accurate and precise method for measuring range of motion of the shoulder and elbow? J Orthop Sci 2018;23:310-5. https://doi.org/10.1016/j.jos.2017.11.016.
28. Sahu D, Shah D, Joshi M, Shaikh S, Gaikwad P, Shyam A. Validation of an on-screen application-based measurement of shoulder range of motion over telehealth medium. J Shoulder Elbow Surg 2022;31:201-8. https://doi.org/10.1016/j.jse.2021.06.017.
29. Scott KL, Skotak CM, Renfree KJ. Remote assessment of Wrist range of motion: inter- and intra-Observer agreement of provider estimation and direct measurement with photographs and tracings. J Hand Surg Am 2019;44:954-65. https://doi.org/10.1016/j.jhsa.2019.05.017.
30. Shimizu H, Saito T, Kouno C, Shimoura K, Kawabe R, Shinohara Y, et al. Validity and reliability of a smartphone application for self-measurement of active shoulder range of motion in a standing position among healthy adults. JSES Int 2022;6:655-9. https://doi.org/10.1016/j.jseint.2022.04.005.
31. Shin SH, Ro du H, Lee OS, Oh JH, Kim SH. Within-day reliability of shoulder range of motion measurement with a smartphone. Man Ther 2012;17:298-304. https://doi.org/10.1016/j.math.2012.02.010.
32. Surgeons AAOO. Joint motion: Method of measuring and recording. American Academy of Orthopaedic Surgeons; 1967.
33. Terwee CB, de Winter AF, Scholten RJ, Jans MP, Devillé W, van Schaardenburg D, et al. Interobserver reproducibility of the visual estimation of range of motion of the shoulder. Arch Phys Med Rehabil 2005;86:1356-61. https://doi.org/10.1016/j.apmr.2004.12.031.
34. Triffitt PD, Wildin C, Hajioff D. The reproducibility of measurement of shoulder movement. Acta Orthop Scand 1999;70:322-4.
35. Wang G, Fiedler AK, Warth RJ, Bailey L, Shupe PG, Gregory JM. Reliability and accuracy of telemedicine-based shoulder examinations. J Shoulder Elbow Surg 2022;31:e369-75. https://doi.org/10.1016/j.jse.2022.04.005.
36. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005;19:231-40. https://doi.org/10.1519/15184.1.
37. Werner BC, Holzgrefe RE, Griffin JW, Lyons ML, Cosgrove CT, Hart JM, et al. Validation of an innovative method of shoulder range-of-motion measurement using a smartphone clinometer application. J Shoulder Elbow Surg 2014;23:e275-82. https://doi.org/10.1016/j.jse.2014.02.030.