

Original Article

Reliability of ultrasound measurement of glenohumeral instability

Catarina Malmberg^{1*}, Kristine Rask Andreassen^{1*}, Jesper Bencke², Birgitte Hougs Kjær³, Per Hölmich¹ & Kristoffer Weisskirchner Barfod⁴

1) Sports Orthopedic Research Center – Copenhagen (SORC-C), Department of Orthopedic Surgery, Copenhagen University Hospital – Amager and Hvidovre Hospital, 2) Human Movement Analysis Laboratory, Department of Orthopedic Surgery, Copenhagen University Hospital – Amager and Hvidovre Hospital, 3) Department of Physical and Occupational Therapy, Institute of Sports Medicine Copenhagen (ISMC), Copenhagen University Hospital – Bispebjerg and Frederiksberg Hospital, 4) Department of Orthopedic Surgery, Institute of Sports Medicine Copenhagen (ISMC), Copenhagen University Hospital – Bispebjerg and Frederiksberg Hospital, Denmark

Dan Med J 2025;72(11):A11240835. doi: 10.61409/A11240835

ABSTRACT

INTRODUCTION. Ultrasound (US) is an accessible and non-invasive method for assessing pathology in the musculoskeletal system. Its application may include assessment of glenohumeral (GH) instability. Whereas GH instability is typically assessed in symptomatic patients, US measurements of anterior-posterior (AP) GH translation have shown good to excellent reliability in healthy populations. The aim of this study was to test the intra- and interrater reliability of measurements of AP GH translation in a clinical setting.

METHODS. Two clinicians performed measures of AP GH translation in patients with anterior shoulder instability. Measurements were conducted from both an anterior and a posterior view. The primary outcome was the intraclass correlation coefficient (ICC(2,1)).

RESULTS. Twenty-three patients were included. The intrarater ICCs ranged from 0.70 to 0.76. The interrater ICCs ranged from 0.20 to 0.49. The most reliable test was a simulated “Load and Shift” test. The minimal detectable changes ranged from 2.16 to 3.42 mm for the same rater, and from 2.71 to 4.78 mm between raters.

CONCLUSIONS. When performed in a clinical setting by raters with limited US experience, the US measurements of AP GH translation in patients with anterior shoulder instability demonstrated moderate to good intrarater reliability and poor interrater reliability. These findings suggest that while US may have potential for clinical use, its reliability may be limited when performed by novice raters.

FUNDING. Amager & Hvidovre Hospitals Forskningspulje.

TRIAL REGISTRATION. ClinicalTrials.gov (ID: NCT05250388).

The glenohumeral (GH) joint is highly mobile and prone to dislocations, affecting nearly 2% of the general population, with many developing recurrent instability [1, 2]. Clinical laxity tests may reveal instability, but are generally examiner-dependent [3]. Objective measurement provides valuable biomechanical insights that can enhance diagnostics, guide rehabilitation and evaluate treatment outcomes. While static imaging modalities like conventional magnetic resonance imaging (MRI) are excellent for assessing soft tissue structures, they do not capture real-time joint movement, which is crucial for understanding functional instability. Dynamic imaging methods like fluoroscopy, advanced dynamic computed tomography (CT) or perioperative arthroscopy offer

detailed assessment, but are limited by radiation exposure or invasiveness [4, 5]. Ultrasound (US) offers an accessible, cost-effective and non-invasive alternative for dynamic GH joint assessment.

Studies suggest increased GH translation in unstable shoulders, but the correlation between translation changes, clinical instability and patient outcomes remains unclear [6]. Research employing radiography, US and motion capture has demonstrated varied findings. For example, biplanar X-ray showed normalised anterior-posterior (AP) translation after open anterior capsule repair, whereas motion capture analysis of open Latarjet procedures found no significant stabilisation [7, 8].

US measures of GH translation have shown promising validity and reliability in healthy subjects, with one study assessing intra- and interrater reliability in anaesthetised patients with anterior shoulder instability [9-14]. However, its reliability in unanaesthetised patients in a clinical context remains unexplored. This study aimed to fill that gap by simulating a clinical setting where multiple physicians with varying US experience assessed patients. We investigated the reliability of AP GH translation measurements in patients with anterior shoulder instability performed by raters with limited US training. Specifically, we investigated 1) intrarater reliability from repeated measurements by a single rater and 2) interrater reliability between two raters, hypothesising good reliability (intraclass correlation coefficient (ICC) > 0.75).

Methods

Study design and trial period

The study followed the Guidelines for Reporting Reliability and Agreement Studies. It included participants from a prospective cohort study of patients with anterior shoulder instability undergoing arthroscopic Bankart repair (from June 2022 to December 2023) [15]. The study was approved by the Danish Capital Region Ethics Committee (H-21027799) and the Knowledge Center for Data Reviews (P-2021-842).

Training of raters

A training phase was conducted to test the setup, refine the protocol and train the raters. Two medical doctors with minimal prior US experience assessed all participants. Rater A had three years of clinical experience, and Rater B had one year. Both received identical training supervised by a physiotherapist with over ten years of musculoskeletal US experience. Additionally, they practised on each other; five asymptomatic individuals and five patients with shoulder instability.

Study execution

A standardised AP GH translation examination protocol was developed (section 2.4) based on prior protocols and pilot trials [10, 16]. Rater A conducted same-day test-retests for intrarater reliability, whereas Rater B performed additional same-day examinations for interrater reliability. All assessments were performed consecutively, but patients were repositioned, and the machine settings were reset between assessments. Each session included one saved image of relevant landmarks (Table 1) ([Supplementary Figures 1-2](#)).

TABLE 1 Ultrasound examination protocol for measurements of anterior-posterior glenohumeral translation.

Step	Details
Rater position	Rater stands behind the patient with the scanner within reach, allowing for adjustments to gain, depth and focus Images are labelled with preset annotations
<i>Posterior view</i>	
Patient position	Patient sits upright on an armless chair, feet flat, shoulder in a neutral position along the body, elbow at 90°, wrist in neutral An assistant aids in maintaining the position
Probe placement & baseline image	Transducer is held parallel to scapular spine, slid towards lateral edge of infraspinatus tendon Humeral head and glenoid are identified; baseline image captured
Isometric force test: flexion	Probe kept in position from baseline image Assistant holds dynamometer at anterior distal upper arm; patient applies 40 N of shoulder flexion force; image captured
Passive laxity test: load & shift	Probe kept in position from baseline image; assistant applies 60 N anterior force at proximal 1/3 of upper arm with dynamometer, image captured
<i>Anterior view</i>	
Patient position	Patient sits upright with shoulder at 90° abduction, arm resting on table, elbow at 90°, wrist neutral An assistant aids in maintaining the position
Probe placement & baseline image	Transducer placed between superior and anterior deltoid parts, in scapular plane Anterior humeral head and coracoid process landmarks located; baseline image captured
Isometric force test: external rotation	Probe kept in position from baseline image; assistant placed dynamometer at distal forearm Patient performs external rotation to 40 N; image captured
Passive laxity test: apprehension test	Probe kept in position from baseline image; assistant applies 60 N force at proximal upper arm; image captured

Blinding

Intrarater

To prevent expectation bias, both image series were captured and stored before measurement. Blinding to prior results was not possible.

Interrater

The examination order was randomised. Each rater independently measured translation on their own images in separate rooms, without discussion.

Ultrasound examination protocol

The AP translation was measured using two probe positions (Table 1) on a Hitachi Arrieta V70 scanner (v00-5.3.0) with a 10 MHz linear probe, mechanical index 0.8 and soft tissue thermal index < 0.4. A custom-preset set was used: depth 30 mm, focus 25 mm, gain (B-mode) 70 dB. The mean of two measurements was analysed. Posterior

measurements followed Rathi et al., and anterior measurements were adapted from Takeuchi et al. [10, 16]. The difference between measurements of distances at rest and under force represented the AP GH translation (Table 1).

Outcomes

The primary outcome was ICC (2,1). Additional outcomes included the standard error of measurement (SEM), minimal detectable change (MDC) and Bland-Altman plots.

Other variables

The following patient demographics were recorded: age, gender, height and weight, limb dominance and affected side.

Study population and eligibility criteria

The eligibility criteria included age 18-40 years, unilateral anterior shoulder instability with radiographically confirmed or reduced dislocation, scheduled arthroscopic Bankart repair, protocol adherence and Danish language proficiency. Written informed consent was obtained. The exclusion criteria were other shoulder pathology, pregnancy or severe illness (American Society of Anesthesiologists (ASA) score ≥ 3).

Recruitment

Patients were recruited from five sports orthopaedic clinics (three public university hospitals; two private hospitals). Eligible patients were identified in outpatient clinics, provided with written study information, and asked to consent to being contacted. Final inclusion and all study activities occurred at the Department of Orthopaedic Surgery, Copenhagen University Hospital – Hvidovre Hospital.

Patient involvement

Pilot trials involving patients with anterior shoulder instability helped shape the design of the examination protocol and outcomes.

Statistical analysis

Sample size

This study, which formed part of a larger cohort study, did not have a pre-hoc sample size estimation. Instead, 23 patients were consecutively included over 18 months. A post-hoc power calculation (R v4.3.0, ICC.Sample.Size package) showed a power of 0.63 to detect an ICC of 0.75, with a minimal ICC of 0.5 and alpha set at 0.05.

Reliability analyses

Reliability was assessed using ICC(2,1) (primary outcome), SEM, MDC, and Bland-Altman plots. The ICC(2,1) model used a two-way random effect, absolute agreement and two measurements ($k = 2$) [17]. ICC values were classified as suggested by Koo et al.: < 0.5 (poor), 0.5-0.74 (moderate), 0.75-0.89 (good), > 0.90 (excellent) [17]. Descriptive statistics included mean ($\pm$ standard deviation (SD)), median (range) and percentages (95% confidence interval). All analyses and figures were created in R (v4.3.0).

Trial registration: ClinicalTrials.gov (ID: NCT05250388).

Results

Patients

Thirty-one patients were recruited, with two excluded due to re-diagnosis of posterior instability and six

withdrawing their consent. Thus, 23 patients participated. Patient data are shown in **Table 2**. No adverse events occurred during or after the experiments.

TABLE 2 Demographic data.

Patients, n	23
Female/male sex ratio, n (%)	6/17 (26.1/73.9)
Age, mean ($\pm$ SD), yrs	28.5 ($\pm$ 6.6)
Height, mean ($\pm$ SD), cm	176.8 ($\pm$ 7.6)
Weight, mean ($\pm$ SD), kg	76.3 ($\pm$ 10.5)
BMI, mean ($\pm$ SD)	24.5 ($\pm$ 3.8)
Dominant side, right/left/both, n (%)	20/2/1 (87.0/8.7/4.3)
Injured side, right/left/both, n (%)	12/11/0 (52.2/47.8/0.0)

SD = standard deviation.

Intrarater reliability

Results are shown in **Table 3**. The Bland-Altman plots for intrarater translation measurements showed no systematic bias, with mean differences below 0.42 mm. However, the limits of agreement were wide, ranging from -3.65 mm to 3.32 mm ([Supplementary figure 3](#)), limiting measurement precision.

TABLE 3 Intrarater reliability.

	Agreement and reliability of measures				Absolute measures, mean, mm ($\pm$ SD)*	
	ICC(2,1) ^a (95% CI)	SEM, mm	MDC, mm (%MDC)	95% LOA, mm	Test 1	Test 2
<i>Posterior view</i>						
Flexion	0.70 (0.4-0.86)	1.07	2.98 (720.1)	-3.37-2.53	-0.62 ($\pm$ 2.06)	-0.20 ($\pm$ 1.86)
Load and shift	0.76 (0.51-0.89)	1.24	3.42 (114.8)	-3.65-3.32	-3.07 ($\pm$ 2.71)	-2.90 ($\pm$ 2.30)
<i>Anterior view</i>						
External rotation	0.71 (0.42-0.86)	0.93	2.59 (115.3)	-2.74-2.50	-2.30 ($\pm$ 1.80)	-2.18 ($\pm$ 1.64)
Apprehension force	0.72 (0.44-0.87)	0.78	2.16 (2066.5)	-2.33-2.13	-0.16 ($\pm$ 1.16)	-0.05 ($\pm$ 1.76)

%MDC = MDC expressed as % of measurement mean; CI = confidence interval; ICC = intraclass correlation coefficient; MDC = minimum detectable change; LOA = limit of agreement; SD = standard deviation; SEM = standard error of measurement.

*) $p < 0.05$ between the measures of Rater A's test 1 and test 2.

a) < 0.50 : poor reliability; 0.50-0.74: moderate reliability; 0.75-0.89: good reliability; > 0.90 : excellent reliability [17].

Interrater reliability

Results are shown in **Table 4**. The Bland-Altman plots of the interrater translation measurements revealed no systematic bias, with mean differences below 1.3 mm. Translation measured in the Apprehension Force test was positively skewed, with rater B recording higher values. The limits of agreement ranged from -4.58 mm to 5.44 mm ([Supplementary Figure 3](#)), reducing measurement precision.

TABLE 4 Interrater reliability.

	Agreement and reliability of measures				Absolute measures, mean, mm ($\pm$ SD)*	
	ICC(2,1) ^a (95% CI)	SEM, mm	MDC, mm (%MDC)	95% LOA, mm	Rater A	Rater B
<i>Posterior view</i>						
Flexion	0.36 (-0.01-0.66)	1.63	4.52 (389.5)	♀-3.30-5.44	-0.62 ($\pm$ 2.06)	-1.69 ($\pm$ 2.02)
Load and shift	0.49 (0.10-0.75)	1.72	4.78 (149.3)	-4.58-5.12	-3.07 ($\pm$ 2.71)	-3.34 ($\pm$ 2.12)
<i>Anterior view</i>						
External rotation	0.20 (-0.24-0.57)	1.48	4.10 (186.3)	♀ -4.33-3.92	-2.30 ($\pm$ 1.80)	-2.10 ($\pm$ 1.51)
Apprehension force	0.36 (-0.09-0.69)	0.98	2.71 (339.6)	-0.99-3.56	-0.16 ($\pm$ 1.16)	-1.44 ($\pm$ 1.29)

%MDC = MDC expressed as a percent of measurement mean; CI = confidence interval; ICC = intraclass correlation coefficient; LOA = limit of agreement; MDC = minimum detectable change; SD = standard deviation; SEM = standard error of measurement.

*) $p < 0.05$ between the measures of Rater A and Rater B.

a) < 0.50 : poor reliability; $0.50-0.74$: moderate reliability; $0.75-0.89$: good reliability; > 0.90 : excellent reliability [17].

Discussion

This study found moderate to good intrarater reliability and poor interrater reliability of US measurements of AP GH translations in patients with anterior shoulder instability, as assessed from ICC. The MDC ranged from 2.71 to 4.78 mm for different raters and from 2.16 to 3.42 mm for the same rater, with a percentual MDC (%MDC) of 114.8-2,066.5%. The simulated “Load and Shift” test, assessing the joint from a posterior view, was the most reliable overall, showing the highest ICC values and the lowest %MDC. While there are no universally defined thresholds for acceptable %MDC in GH translation, values above 100% suggest that the measurement error exceeded the measurement itself, making it difficult to distinguish actual changes from measurement variability. Additionally, our lowest observed %MDC of 114.8% raises concerns regarding the clinical applicability of the current protocol.

Overall, the results indicate that the proposed US protocol lacks reliability when used by operators with limited US experience in a clinical setting. For future research, a reliability study involving more experienced clinicians is recommended.

Our results differ from previous studies that reported good to excellent intrarater reliability (ICC: 0.81-0.998) and poor to excellent interrater reliability (ICC: 0.31-0.98), likely due to differences in patient populations, such as healthy subjects, anaesthetised patients or cadavers [10, 14, 16].

The lower ICC values in this study were expected, as patient variability generally exceeds that of healthy subjects. Both raters in this study had limited US experience. In contrast, both studies mentioned above involved radiologists or orthopaedic surgeons with extensive US experience, which could explain the higher ICC [14, 16]. Additionally, the study by Inoue et al. involved only one rater acquiring the US images, which may have reduced variability in image acquisition technique. Furthermore, novice US raters may exhibit greater variability, leading to lower ICC values.

Even slight variations in probe positioning and angulation could influence measurements. This may have led to inaccuracies in capturing the true anterior or posterior translation, as subtle probe misalignment may cause the US beam to assess a slightly oblique or off-axis plane rather than a strictly AP direction.

ICC(2,1) was used to assess interrater reliability, assuming that raters were randomly selected from a larger population and evaluating the reliability of a single measurement per rater. Whereas ICC(2,k) could provide higher reliability estimates by averaging multiple ratings, our design involved one assessment per rater (although based on the average of two measurements), making ICC(2,1) the most appropriate choice [18]. For intrarater reliability, a fixed-effects model like ICC(3,k) with $k = 2$ could have been used. This would potentially

have yielded higher estimates by eliminating interrater variability [17]. Recognising the limitations of statistical methods, we applied multiple statistical approaches (ICC, SEM, MDC and Bland-Altman plots) to ensure consistency and minimise interpretation uncertainties [19]. In our case, the results remained consistent across methods, reinforcing the robustness of our findings.

Rater A registered mean anterior translations ranging from 0.05 mm (SD: ± 1.76) to 3.07 mm (SD: ± 2.71), whereas Rater B measured 1.44 mm (SD: ± 1.29) to 3.34 mm (SD: ± 2.12), with the highest values observed during the “Load and Shift” test. These values were lower than those reported by Inoue et al. at 0° of abduction in anaesthetised patients (5.29 mm anterior translation) [14]. Krarup et al. reported a mean anterior translation of 4.9 mm in an adducted arm under a 90 N force, which is more comparable to our findings using a 60 N force [12]. Krarup et al. further found a mean anterior translation of 2.1 mm in the healthy shoulders. Variability in study protocols, including force, positioning and patient conditions, complicates direct comparisons between studies. As a result, no established threshold for abnormal AP GH translation has been defined [6].

Capturing all image series before measurements limited expectation bias, and Raters A and B were blinded to each other’s results. However, the rater was not blinded to their own results for intrarater reliability. The limited sample size of the study affects statistical power and may lead to type II error, and a larger sample would have made the results more robust.

A major strength of this study was its resemblance to clinical practice. The study protocol was simple and potentially transferable to the clinical setting, e.g., at the outpatient clinic, though requiring an assistant could limit its clinical use.

Manual clinical tests remain essential for diagnosing shoulder instability, but examiner experience often limits their reliability. Dynamic imaging methods offer real-time objective assessments of GH translation and may enhance diagnostic accuracy and treatment evaluation. We acknowledge that no single imaging modality currently provides an optimal solution for tracking GH joint motion [6]. While stress-loaded cross-sectional imaging, such as dynamic MRI, offers the advantage of non-irradiating 3D visualisation, it is limited by scan time, availability and challenges in capturing high-speed movements. Although recent advancements in MRI protocols have shown potential, their reliability and feasibility require further validation [6]. Radiostereometric analysis (RSA) provides accurate 3D kinematic measurements, but is constrained by radiation exposure and requires specialised equipment [20]. US is widely available and non-invasive, allowing for real-time dynamic assessment without radiation. While US lacks the depth resolution of MRI and the precision of RSA, its ability to provide functional insights during movement in a clinical setting makes it a valuable tool for assessing GH translations. However, its clinical applicability for GH translation has yet to be fully established, particularly regarding validity and reproducibility across different examiners and patient populations. Each method has inherent advantages and limitations, and the choice of imaging technique should be guided by the specific clinical or research question at hand.

Conclusions

Intrarater reliability was moderate to good, whereas interrater reliability was poor among raters with limited US experience for measuring AP GH translation in patients with anterior shoulder instability. The simulated “Load and Shift” test demonstrated the highest reliability. However, the large variability of the measurements and high %MDC values cast doubt on the clinical applicability of this US protocol when performed by inexperienced sonographers.

Correspondence *Catarina Malmberg*. E-mail: catarina.anna.evelina.malmberg.02@regionh.dk

*) Shared first authorship

Accepted 15 August 2025

Published 15 October 2025

Conflicts of interest KWB reports financial support from or interest in Company DJO, LLC. KRA reports financial support from or interest in Company Amager and Hvidovre Hospitals Forskningspulje. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. These are available together with the article at ugeskriftet.dk/dmj

References can be found with the article at ugeskriftet.dk/dmj

Cite this as Dan Med J 2025;72(11):A11240835

doi 10.61409/A11240835

Open Access under Creative Commons License [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Supplementary material: [a11240835-supplementary.PDF](#)

REFERENCES

1. Hovelius L. Incidence of shoulder dislocation in Sweden. Clin Orthop Relat Res. 1982;166:127-131
2. Hovelius L, Saeboe M. Neer Award 2008: arthropathy after primary anterior shoulder dislocation - 223 shoulders prospectively followed up for twenty-five years. J Shoulder Elbow Surg. 2009;18(3):339-347. <https://doi.org/10.1016/j.jse.2008.11.004>
3. Hegedus EJ, Goode AP, Cook CE, et al. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. Br J Sports Med. 2012;46(14):964-978. <https://doi.org/10.1136/bjsports-2012-091066>
4. Di Giacomo G, Itoi E, Burkhart SS. Evolving concept of bipolar bone loss and the Hill-Sachs lesion: From "engaging/non-engaging" lesion to "on-track/off-track" lesion. Arthroscopy. 2014;30(1):90-98. <https://doi.org/10.1016/j.arthro.2013.10.004>
5. Jahnke AH Jr, Greis PE, Hawkins RJ. Arthroscopic evaluation and treatment of shoulder instability. Orthop Clin North Am. 1995;26(4):613-630
6. Malmberg C, Andreassen KR, Bencke J, et al. Anterior-posterior glenohumeral translation in shoulders with traumatic anterior instability: a systematic review of the literature. JSES Rev Rep Tech. 2023;3(4):477-493. <https://doi.org/10.1016/j.xrrt.2023.07.002>
7. Paletta GA Jr, Warner JJ, Warren RF, et al. Shoulder kinematics with two-plane x-ray evaluation in patients with anterior instability or rotator cuff tearing. J Shoulder Elbow Surg. 1997;6(6):516-527. [https://doi.org/10.1016/S1058-2746\(97\)90084-7](https://doi.org/10.1016/S1058-2746(97)90084-7)
8. Lädermann A, Denard PJ, Tirefort J, et al. Does surgery for instability of the shoulder truly stabilize the glenohumeral joint? A prospective comparative cohort study. Medicine (Baltimore). 2016;95(31):e4369. <https://doi.org/10.1097/MD.0000000000004369>
9. Rath S, Taylor NF, Green RA. The effect of in vivo rotator cuff muscle contraction on glenohumeral joint translation: an ultrasonographic and electromyographic study. J Biomech. 2016;49(16):3840-3847. <https://doi.org/10.1016/j.jbiomech.2016.10.014>
10. Takeuchi S, Chan CK, Hattori S, et al. An improved quantitative ultrasonographic technique could assess anterior translation of the glenohumeral joint accurately and reliably. Knee Surg Sports Traumatol Arthrosc. 2021;29(8):2595-2605. <https://doi.org/10.1007/s00167-021-06459-1>
11. Joseph LH, Hussain RI, Pirunsan U, et al. Clinical evaluation of the anterior translation of glenohumeral joint using ultrasonography: an intra- and inter-rater reliability study. Acta Orthop Traumatol Turc. 2014;48(2):169-174. <https://doi.org/10.3944/AOTT.2014.3184>
12. Krarup AL, Court-Payen M, Skjoldbye B, et al. Ultrasonic measurement of the anterior translation in the shoulder joint. J

- Shoulder Elbow Surg. 1999;8(2):136-141. [https://doi.org/10.1016/S1058-2746\(99\)90006-X](https://doi.org/10.1016/S1058-2746(99)90006-X)
13. Court-Payen M, Krarup AL, Skjoldbye B, et al. Real-time sonography of anterior translation of the shoulder: an anterior approach. *Eur J Ultrasound*. 1995;2(4):283-287. [https://doi.org/10.1016/0929-8266\(95\)00114-7](https://doi.org/10.1016/0929-8266(95)00114-7)
 14. Inoue J, Takenaga T, Tsuchiya A, et al. Ultrasound assessment of anterior humeral head translation in patients with anterior shoulder instability: correlation with demographic, radiographic, and clinical data. *Orthop J Sports Med*. 2022;10(7):23259671221101924. <https://doi.org/10.1177/23259671221101924>
 15. Kottner J, Audigé L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64(1):96-106. <https://doi.org/10.1016/j.jclinepi.2010.03.002>
 16. Rath S, Taylor NF, Gee J, et al. Measurement of glenohumeral joint translation using real-time ultrasound imaging: a physiotherapist and sonographer intra-rater and inter-rater reliability study. *Man Ther*. 2016;26:110-116. <https://doi.org/10.1016/j.math.2016.08.001>
 17. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
 18. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428. <https://doi.org/10.1037//0033-2909.86.2.420>
 19. de Vet HCW, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59(10):1033-1039. <https://doi.org/10.1016/j.jclinepi.2005.10.015>
 20. Kipp JO Petersen ET, Falstie-Jensen T, et al. Glenohumeral joint kinematics during apprehension-relocation test in patients with anterior shoulder instability and glenoid bone loss. *Bone Joint J*. 2024;106-B(10):1133-1140. <https://doi.org/10.1302/0301-620X.106B10.BJJ-2024-0419.R1>