Evidence-Based Practice Group Answers to Clinical Questions

Validity and Reliability of Electrogoniometer in Musculoskeletal Injury/Disorder ROM Assessment

By

WorkSafeBC Evidence-Based Practice Group

Dr. Craig Martin

Manager, Clinical Services

Chair, Evidence-Based Practice Group

May 2018

WorkSafeBC
WORKING TO MAKE A DIFFERENCE

Clinical Services – Worker and Employer Services

About this report

Validity and Reliability of Electrogoniometer in Musculoskeletal Injury/Disorder ROM Assessment

Published: May 2018

About the Evidence-Based Practice Group

The Evidence-Based Practice Group was established to address the many medical and policy issues that WorkSafeBC officers deal with on a regular basis. Members apply established techniques of critical appraisal and evidence-based review of topics solicited from both WorkSafeBC staff and other interested parties such as surgeons, medical specialists, and rehabilitation providers.

Suggested Citation

WorkSafeBC Evidence-Based Practice Group, Martin CW. Validity and Reliability of Electrogoniometer in Musculoskeletal Injury/Disorder ROM Assessment
Richmond, BC: WorksafeBC Evidence-Based Practice Group; May 2018

Contact Information

Evidence-Based Practice Group
WorkSafeBC
PO Box 5350 Stn Terminal
Vancouver BC V6B 5L5

- Email • craig.martin@worksafebc.com
- Phone • 604 279-7417
- Toll-free • 1 888 967-5377 ext 7417

Objective

To examine the validity/reliability of electrogoniometers in the assessment of range of motion in musculoskeletal injuries/disorders

Background

Assessment of range of motion (ROM) to determine the movement capacity and limitation of joints is an essential part of the clinical musculoskeletal system examination. Both the joint and surrounding tissues are studied. Active (unassisted, voluntary) and passive (assisted by the examiner/rater/tester or medical equipment) ROM assessments are useful in diagnosing and managing various musculoskeletal conditions, including work-related musculoskeletal injuries. However, when performing ROM assessments on patients with unknown etiology, especially during passive ROM measurements, caution should be exercised in order to avoid potential physical complications.[1-3] Even if a 0-360 degree measurement would be more valuable in recording ROM for some joints (e.g., shoulder); most of the time a 0-180 degree system is used where 0-degree starting point is the anatomical position of the joint.

Range of motion can be assessed using different methods (e.g., clinical examination/visual inspection, screening the adjacent and/or contralateral joints, using rulers/tape measures, goniometers/inclinometers, photography, and radiological techniques such as x-rays, MRIs). During goniometric assessment the starting point for ROM is determined by the rater (knowledge/skill/experience dependent), whereas during inclinometric assessment it is determined by gravity. Typically, x-rays[4-7] or goniometers[8, 9] are cited as the gold standard measurement tools for ROM.

What is goniometer?

Goniometer is a type of instrument used to assess joint ROM in different joint planes by measuring angles between different joint structures (i.e., bones). Goniometry dates back to the ancient Greece (gonia and metron meaning 'angle' and 'measure', respectively) and have been used in medicine since early 1900s.[10] A classic goniometer consists of two arms, a protractor and an axis. The arms are usually 12-inch long and sometimes difficult to pin down to the exact landmark being studied. The landmark, measurement starting point, is determined by the examiner. Some goniometers are able to measure both flexion and extension of the joint

without repositioning the goniometer.[3] In occupational context, goniometers may be used to assess ROM after the initial injury and to document subsequent progress during the rehabilitation process and for determining the level of permanent disability.

In the literature various adjectives are used to define goniometers (e.g., standard, universal, functional, manual, digital, electro-, fluid-, telescopic-armed, protractor based, smart-phone based, mobile device based). Each of these goniometers may have different features. For this review, we will be focusing on electrogoniometers, which are more commonly used in ROM assessments in occupational settings.

Electrogoniometers are useful in the functional evaluation of different joints. Their light weight, portability, low fragility, flexibility to wear, relative lower cost, data recoding capacity (even for longer periods of time) are advantages.[11, 12] They measure ROM in a continuous manner and are useful in ergonomic assessments and in evaluating repetitive movements. Different technologies (e.g., potentiometers, strain gauges, accelerometers) can be used to implement an electrogonimeter.[13] In occupational health, electrogoniometers are used in evaluating both upper and lower extremity movement in workers from various industries (e.g., fish processing workers, tree nursery workers, office workers and cleaners, keyboard operators, carpenters, carpe/floor layers, farmers, miners, soldiers, elite dancers, athletes). [12] One advantage of electrogoniometers is their more accurate numerical presentation of angulation in decimal places compared to 1 degree increments available with universal standard goniometer. Also, they can be used simultaneously with dynamometry and electromyography in clinical settings, which may help produce more accurate ROM records for clinical research.[14] However, a careful positioning and calibration of the goniometer (according to task) is required to avoid errors.[3] Characteristics of the subjects studied (e.g., skin flexibility, structure of bones, fat and muscle build) may also affect the results.[15] Compared to their intraexaminer reliability, interexaminer reliability of electrogoniometers seem to be lower.[16] One other disadvantage of electrogoniometers is cross-talk, which can be inherent or resulting from rotation, flexion, extension or deviation. The inherent sensor crosstalk is based on how sensitive elements ("strain gauges") are housed inside the device. Errors due to either inherent cross-talk or spring torsion [when testing flexion/extension, rotation, deviation, as appropriate], both increase during movements with greater ROM.[11] Although, not confirmed by their own study, Bronner et al. mention that there are reports on decreasing accuracy during 'motion extremes' when studying upper limb motion with electrogoniometers.[12] Despite the efforts to standardize goniometers and increase their reliability, scores are affected by varying models, material characteristics (e.g., stiffness) and varying marking processes of different goniometric devices.[17]

ROM Assessment

According to the American Medical Association (AMA) Guides, impairment rating is "a consensus-derived percentage estimate of loss of activity, which reflects severity of impairment for a given health condition, and the degree of associated limitations in terms of activities of daily living (ADLs)"[18] A thorough clinical assessment to estimate impairment employs valid and reliable measurement tools (e.g., questionnaires, laboratory tests, equipment). In the case of musculoskeletal conditions, the outcome of interest is often range of motion (ROM) of the joint, and various factors should be considered during ROM assessment. Goniometers, as indicators of joint movement capacity, are widely used for this purpose. A preferable goniometer is expected to convey a measurement that represents the actual joint angle or total ROM. During follow-up, it is important to know if the observed change in ROM is due to a true change in patient's clinical state or result of a measurement error inflicted by the goniometer. Therefore, data enabling quantify the level of measurement error should be included in studies.[19] The content validity of a goniometer would incorporate knowledge, experience and skills of the rater (practitioner performing the measurement), because the visual inspection and palpation of the anatomical landmarks, as well as the correct alignment of the goniometer may be crucial for a valid goniometric assessment. Certain structural characteristics of joints may make positioning and stabilization of the goniometer difficult. When standard/manual goniometers are used there is usually no blinding, and examiners may introduce their individual biases during reading/recording of the measurements. In general, when measurements are repeated by the same examiner/rater using the same goniometer the measurement error tends to be small; whereas when the goniometer or the rater is changed measurement errors may become significant.

There are a few problematic issues in the literature with regards to the utilization of goniometers in ROM assessment.

- Inconsistent and ambiguous definitions/terminology for musculoskeletal disablement
- Questionable content and predictive validity of goniometers
- Questionable reliability of goniometers
- Lack of internal consistency in goniometry studies
- Inadequate evidence-base (limited number of high quality studies)
- Disability rates not reflecting the actual or perceived loss of musculoskeletal function

Validity and Reliability

To use a measurement tool (e.g., test, questionnaire, equipment) it is preferred that the scores/results it generates are both valid and reliable. If it is either or, the choice on a reliability or validity study should be determined based on the clinical context that is being studied. More often, existing studies on ROM assessment focus on reliability, rather than validity. However, reporting on validity is important; as such, a distorted instrument can persistently produce reproducible, but biased (e.g., systematically erroneous) results. On the other hand, it should also be recognized that validity measured in a certain study population may not necessarily be extrapolated to other populations.

Validity

The AMA Guides to the Evaluation of Permanent Impairment defines validity as "the extent to which an instrument or test actually measures what it is intended to measure", and accuracy as "the quality of being correct or near to the truth, especially the degree to which a measurement, calculation, or estimate conforms to the true value".[18] Validity of a test/study can be improved when systematic errors (leading to bias) are avoided.[20]

- **Construct validity**
  Construct validity is the extent that a measurement instrument can be considered measuring the studied 'construct' (e.g., accompanied conceptual elements of the same phenomenon), capturing the studied concept, and can be used for making related inferences. For example, a goniometric ROM measurement being used to make inferences about disability.
- **Face validity**
  Face validity is about plausibility. In other words, it basically informs us that it is plausible to use this measurement instrument to measure what was originally intended to be measured.
- **Content validity**
  Content validity tells us about the extent of the coverage (depth, rigor, detail) accomplished by a measurement instrument on the studied concept/domain. However, 'content validity' is a subjective call, which might not only reflect the properties of the measurement instrument, but also the knowledge/skills/interpretation of the examiner/rater.
- **Criterion validity**
  Criterion validity of a measurement instrument realized by comparing the measurements from this instrument to the measurements from an established/recognized/gold-standard instrument (i.e., criterion). Criterion validity can be objectively tested with statistical methods.

- Concurrent validity is a type of criterion validity and is the extent that the measurements from the instrument being tested (e.g., test, questionnaire, equipment) and the ones from the criterion taken approximately at the same time are in line with each other.
- Predictive validity is another type of criterion validity, which is the extent that the current scores from a measurement instrument holds when compared with the scores from a criterion (another, established measurement instrument) in the future.
- Convergent validity is the extent the measurements from the studied instrument are similar to the ones (converging) from another instrument (criterion) to which they were theoretically expected to be similar.
- Divergent (discriminant) validity is the extent the measurements from the studied instrument are different from the ones (diverging) from another instrument (criterion) from which they are theoretically expected to be different.

Radiography has been regarded as the gold-standard to test the criterion validity of goniometers.[4, 5, 7, 21] Despite some variations in the observed correlation coefficients and intraclass correlation coefficients (ICCs) depending on the joint being studied, in general, concurrent validity of radiology and goniometers was found to be high.[2]

Sometimes, validity is used as a general term to define the methodological properties of a study. For example, 'internal validity' may refer to the proper conduct (with no major methodological flaws) of a study and 'external validity' may refer to the true representation of the reference population by the sample studied, hence allowing generalizability of the findings beyond the studied sample.

Reliability

In mathematics, reliability represents "a ratio of true variance over true variance plus error variance".[22] The AMA Guides to the Evaluation of Permanent Impairment defines reliability as "the extent to which a test or measurement yields consistent results when repeated",[18] pointing to the reproducibility of measurement results by different people, at different times and places, given similar conditions. In a way, reliability represents consistency, stability of the findings when measurements are undertaken in different, but similar conditions. Lack of reliability may obscure the relationship sought between different variables.[20] Precision is defined as "the smallest unit of change a measuring instrument can distinguish, or the number of digits used to express the measurement".[18] Since reliability should reflect both "degree of correlation and agreement between measurements", intraclass correlation coefficient (ICC) is an appropriate

index to measure it.[22] However, there are different types of ICC calculations and it is important that the one appropriate for the current data/study design is chosen and also declared in publications to allow for unflawed interpretation.

- Test-retest reliability (repeatability)
  Test-retest reliability is determined by taking repetitive measurements on the same subject, under the same conditions and at different points in time. The extent of variation is computed. Usually, when measuring test-retest reliability the examiner/rater effect is negligible or does not exist (e.g., a self-report questionnaire).
- Intrarater reliability
  Intrarater reliability aims to capture the variation when the measurements are undertaken by the same rater using the same data, across multiple points in time. Usually, intrarater reliability would be checked by calculating intraclass correlation coefficients (ICC) and associated 95% confidence intervals (CI) for comparison across different measurement occasions.
- Interrater reliability
  Interrater reliability measures the "variation between two or more raters who measure the same group of subjects".[22] Measurements are undertaken by different raters, on the same data set, using the same instruments, and at close points in time. In a way, interrater reliability reflects the consistency of measurements across raters (ICC and associated 95% CI would be calculated for comparison across measurement scores by different raters).
- Interinstrument reliability
  Interinstrument reliability aims to identify the variation between the measurements undertaken using different instruments

Validity and Reliability of Electrogoniometer

Gonimeters have been around for a long time. For example, the paper on reliability of ROM measurements by Armstrong[17] includes references from the end of 1920s. However, the goniometer literature does not offer many studies on electrogoniometers and the number of studies that assessed the validity and reliability of electrogoniometers is even scarcer.

Electrogoniometers are usually validated against radiography, which is considered as the gold-standard for ROM assessment by many clinicians and researchers.[4, 5, 7, 21]

While the face validity appears to be met for electrogoniometers in general; it is difficult to establish their content validity. This is because of the nature of the profoundly multifactorial environment they are operated in. For example, the knowledge and skill of the examiner and the structure of the

joint being studied all affect the success of the measurement. For criterion validity, electrogoniometry is often compared with radiographic studies (e.g., x-rays, MRI/CT scan) as the gold standard. There are not many electrogoniometer studies focusing on the specific types of the criterion validity (i.e., concurrent, predictive, convergent, divergent validities). At least one study[23] implied the construct validity of electrogoniometers by emphasizing how goniometric measurements can be used in measuring functional movement of the joints and disability.

Most of the electrogoniometry studies identified for this review reported on the reliability of the equipment. Especially, interrater and intrarater reliability were assessed.[7, 12, 14, 23, 24] Often, the validity tested was 'concurrent validity'.[7, 12] The construct validity of electrogoniometers was assessed in the context of disability and function.[7, 23]

Methods

- The EBPG conducted a systematic literature search on validity and reliability of electrogoniometers, on January 31, 2018.
- We searched the commercial medical literature databases available through the OVID SP platform, which included Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Health Technology Assessment, NHS Economic Evaluation Database, BIOSIS Previews, Embase, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R)
- **Search Strategy**
  We employed a search strategy combining relevant keywords with Boolean operators, AND/OR, as appropriate:

  [(Range of motion OR ROM) AND (Examination OR Assessment OR Analysis OR Testing OR Evaluation OR Measurement)**]**

  AND

  [Goniometry OR Goniometer OR Goniometre OR Electrogoniometer OR Electrogoniometre**]**

  AND

  [Validity OR Reliability OR Precision OR Accuracy**]**

- When searching for articles on 'electrogoniometer', we included the broader term 'goniometer'. With this approach our search was more inclusive and aimed to capture any studies on 'electrogoniometer' that may have not specified the type of the goniometer as a keyword
- There were 725 citations identified with this search strategy
- We limited the search to citations on adult populations (ages <18 to 64>), written in English, and published in the last 10 years. The number of citations were down to 335
- After removal of the duplicates, the number of citations was 260
- Abstracts for these 260 citations were scanned employing the inclusion/exclusion criteria outlined below

- **Inclusion/exclusion criteria** (first round)
    - Include articles
        - on goniometers
        - on study populations of healthy individuals or patients with musculoskeletal injuries/disorders
        - in the formats of systematic reviews, meta-analyses, randomized controlled trials, cohort studies, comparative studies, case-control studies, cross-sectional studies, case series
    - Exclude articles
        - in the formats of single case reports, conference abstracts, narrative reviews, study protocols, letters to editors
        - on cadavers
        - on animal MSDs (musculoskeletal disorders)
        - on pediatric age group MSDs
        - on stroke patients
        - on patients with chronic brain and spinal cord conditions (e.g., ankylosing spondylitis, cerebral palsy)
        - on burn patients
        - on patients using prosthetics
        - solely on fracture patients
        - solely on postoperative surgery patients
        - solely on cost-effectiveness of goniometers

- **Inclusion/exclusion criteria** (second round)
    - Include articles
        - on validity and reliability of electrogoniometers in ROM assessment
    - Exclude articles
        - on types of goniometers other than electrogoniometers

Results

When 260 abstracts from the 335 Ovid-SP search citations were scanned based on the content and our first round of inclusion/exclusion criteria outlined above, 161 of them were excluded.

The remaining 99 abstracts were on range of motion (ROM) measurements of different joints, using various goniometers. Nineteen percent had ROM measurements specifically on knee, 15% on shoulder, 13% on wrist, 12% on hand, 8% on hip, 7% on each of elbow and neck, 5% on spine, 4% on ankle, 2% on foot joints. Thirty two percent of the abstracts reported multiple joint ROM measurements and 7% did not specify the joint studied and used general terms, such as upper or lower extremities.

Out of the 99 abstracts, only 6 (6%) were on validity and/or reliability of electrogoniometry in measuring ROM in musculoskeletal disorder (MSD) patients or healthy subjects. Full text of these six articles were collected for critical appraisal and their characteristics are summarized in Table 1.

Table 1 – Characteristics of the validity/reliability studies selected

Author / Year	Study design / type	Study population	Study sample	Sample size	Joint(s) studied	Measurement tool tested	Examiner/ Rater background	Testing interval	Analysis/ reporting	Results	Conclusion/ Notes
Bashardoust Tajali 2016 [7]	Cross-sectional study - Reliability (intrarater, interrater inter instrument) - Validity (construct)	Outpatients of the Hand and Upper Extremity Center (HULC) at St. Joseph Health Care Center in London Ontario	Patients with limited wrist or hand motions	44 patients -24 Female -20 Male	- Wrist (active ROM) - Index finger (active & passive ROMs) Torques of index finger PIP passive flexion * Quantifying torque applied during ROM assessment is important, as torque applied might affect ROM obtained by different raters	- Two digital electrogoniometers (NK Hand Assessment Laboratory Joint Motion & J-Tech electrogoniometer)	- Physical therapist - Kinesiologist	2-5 days between measurements	- Reliability: Tests of difference & correlation coefficients (e.g., ICCs) were used for intrarater, interrater, inter-instrument comparisons - Construct validity: correlation coefficients were used for comparisons with PRWE & quick DASH	- Bland-Altman plots for mean differences btw comparisons (caught no systematic errors) - Index finger PIP torque applied by the 2 raters during passive ROM (statistically different)	- NK & J-Tech with high reliability coefficients, tight error margins in active wrist ROM and active/passive PIP index flexion - ROM findings contributed to the construct of functional disability - Other finger digits, criterion validity were not studied
Bronner, 2010 [12]	- Validity/ accuracy/ differences in mid- and end-range testing (static test) - intrarater & instrument reliability/ accuracy/ concurrent validity compared to 3D Motion	College student enrolled in a major in dance bachelor degree program	Advanced-level dancers (average 10-year experience)	17 dancers - 10 female - 7 male (written informed consent optained)	- Hip - Knee - Ankle	- Three biaxial electrogoniometers (two SG150 (150 mm) & one SG110 (110 mm) flexible electrogoniometers (Penny and Giles, Biometrics)	?	- Same day Instrument reliability - 2 day interval Intra-rater reliability	Static testing (used SPSS); accuracy was measured by SEM, ICCs were computed to compare (using two-tailed t-tests) electrogoniometers & digital protractor measurements, both at mid-	Static test: electrogoniometer to protractor correlations (r≥0.99, SEM≤3.65°); Dynamic test instrument & intra-rater reliability correlations (r≥0.98 and r≥0.97, SEM≤3.49° and ≤4.48°), concurrent validity correlations	Reliability and validity correlations were high; as well as the dynamic measurement errors (SEM≤6.80°). Authors argue that this high error "may be acceptable for motion studies

btw: between; DASH: Disabilities of the Arm, Shoulder and Hand; ICS: Intraclass Correlation Coefficients; PIP: proximal interphalangeal; PRWE: Patient Rated Pain and Function; ROM: Range of Motion; SEM: standard error of measurement

	analysis (dynamic test)					- digital protractor (Bosch Digital Protractor/ Angle Finder DWM40L) - 5 camera motion analysis system (Vicon)			range and at end-range - Dynamic testing for reliability (ICCs) and validity (SEM), using 3D Motion analysis as the criterion reference	(r≥0.94, SEM≤6.80° hip, knee, ankle combined) to motion analysis	or workplace exposure testing where motion analysis is not available or feasible"
da Silva Camassuti 2015 [14]	- Reliability (intrarater, interrater, interdevice)	Healthy right-handed volunteers tested in university hand and upper limb clinical research laboratory	A sample of right-handed healthy individuals	24 individuals - 12 Female - 12 Male	- Wrist (radial & ulnar deviation, flexion and extension of the right wrist with random, Active ROM)	- universal goniometer (CARCI) - electrogoniometer (MIOTEC)	two examiners with prior training in how to use the devices	- Interrater reliability (3 series of measurements) - Intrarater reliability (one examiner repeated measurement after 7 days)	Intrarater, interrater, interdevice reliability was assessed by computing ICCs & SEMs, and systematic errors were checked for using Bland-Altman plots	(0.70-0.79 to >0.90) - Intrarater reliability: excellent using electrogoniometry and moderate using goniometry - Interinstrument reliability: moderate (when all wrist motions were considered	- Interrater reliability: moderate to excellent - Bland-Altman limits of agreement: poor for interrater assessments of ulnar & radial deviation, and for interdevice agreements of all ROM
Law 2013 [23]	Cross-sectional study - Reliability (intrarater, interrater) - Validity (construct)	Outpatients from the Physiotherapy Department, the United Christian Hospital, Hong Kong	Convenience sample	52 patients -26 patients with neck pain -26 volunteers without neck pain	- Cervical spine AROM (active ROM) (cervical flexion-extension, side flexion, rotation)	ACRON Cervical Goniometer (composed of dual electronic inclinometers	Two physiotherapists (with 5 year clinical experience)	- 7 days between measurements (Intrarater reliability) - two rooms used by two raters during the same measurement session (interrater reliability)	- Used a patient and a control group to test construct validity - ICCs were computed for intrarater and interrater reliability	- intrarater & interrater reliability (ICC) for cervical ROM ranged from 0.75 to 0.92 in control & patient groups - Patient group total cervical AROM was significantly smaller ($p < 0.001$) than the control group AROM (supporting construct validity)	ACRON electronic cervical goniometer was reliable (for 3 cervical mobility planes for both normal subjects and patients) Construct validity was supported as a significant difference in

btw: between; DASH: Disabilities of the Arm, Shoulder and Hand; ICS: Intraclass Correlation Coefficients; PIP: proximal interphalangeal; PRWE: Patient Rated Pain and Function; ROM: Range of Motion; SEM: standard error of measurement

											AROM between control & neck pain groups
Piriyaprasarth, 2008 [24]	Reliability study - interrater - intrarater	Study1 La Trobe University undergraduate and postgraduate students & staff Study2 physiotherapy students (The University of Melbourne, Australia)	Healthy volunteers (convenience sample)	Study1 35 healthy adults (mean age: 31) - 9 male - 26 female Study2 20 healthy adults (mean age: 20) - 1 male - 19 female	Study1 (Both knees) sagittal measurements in supine, sitting and standing positions Study2 (Right knee) detailed tests during neutral standing position & after 10m walking	Penny and Giles Biometrics® (P & GB) twin axis electrogoniometer	Study1: Two physiotherapists with more than 3-year clinical experience & 1-3 month experience with the electrogoniometer Study2: Two physiotherapists; one experienced, the other received 2-hour training on the study electrogoniometer	All tests were done on the same day	Study1: ICCs, SEMs, and limits of agreement for inter-tester and intratester reliability of knee joint were computed across three testing positions (sitting, supine, standing) Study2 SEMs, and limits of agreement for inter-tester and intratester reliability of knee joint were computed for standing and was repeated after 10-meter walk	Study1: Intertester ICCs: 0.58-0.71 (supine), 0.68-0.79 (sitting), 0.5-0.80 (standing). SEMs btw testers ≤3.55; limits of agreement were -12.51°-12.21°. Intratester ICCs: 0.75-0.76 (supine), 0.86-0.87 (sitting), 0.87-0.88 (standing). SEMs for same tester measurements ≤1.7°; limits of agreement: -8.13°-7.90°. Study2: Intertester reliability, (SEMs for standing & after 10-min walk) was 0.5°-3.3°. Limits of agreement: -6.0°-4.9°. Intrarater reliability: (SEMs standing and after 10-min walk) was 1.3°-2.3°. Limits of agreement: -6.7°-2.9°.	Using a standard goniometer attachment protocol during measurement of knee movement by a flexible goniometer in standing, supine and sitting positions may help minimize measurement error and may increase reliability
Zampagni 2008 [25]	Experimental analysis of reliability (for elbow carrying angle)	Master swimmer athletes (retired)	37 healthy (with no symptoms re: shoulder, elbow,	37 healthy adults - 17 men - 20 women	- Elbow	Faro Arm (electrogoniometer)	Orthopedic surgeons	?	- ICCs were computed for intrarater and interrater reliability	- Mean carrying angle was 12.7 ± 3.8° (pointing to great individual variability) - Interrater reliability was good	- High variability in standard deviation for the carrying angle in study sample demonstrated

btw: between; DASH: Disabilities of the Arm, Shoulder and Hand; ICS: Intraclass Correlation Coefficients; PIP: proximal interphalangeal; PRWE: Patient Rated Pain and Function; ROM: Range of Motion; SEM: standard error of measurement

			wrist), former swimmers, 41 to 81 years of age, who gave informed consent	(Total 72 measurements, from right and left arms					- t-test for dependent and independent samples was used for comparisons. Testing was repeated with nonparamentric Mann-Whitney and Wilcoxon tests	(ICC=0.7) and intrarater reliability was excellent (ICC=0.85) - Neither limb side (p= 0.76) nor gender (p= 0.57) differences were significant	individual variability - The authors concluded that carrying angle value and its pathological variations using this method was fast and suitable for clinical applications

btw: between; DASH: Disabilities of the Arm, Shoulder and Hand; ICS: Intraclass Correlation Coefficients; PIP: proximal interphalangeal; PRWE: Patient Rated Pain and Function; ROM: Range of Motion; SEM: standard error of measurement

Reviewed Studies

Reliability and Validity of Electro-Goniometric Range of Motion Measurements in Patients with Hand and Wrist Limitations

(Bashardoust Tajali, 2016)[7]

The authors conducted a study to evaluate two electrogoniometers, NK and J-Tech, in 44 patients with limited wrist/finger motion. They compared both active and passive range of motion (ROM) to test intrarater, interrater and interinstrument reliabilities and construct validity of these electrogoniometers. The testing was conducted using randomized block design. Two experienced raters, a physical therapist and a kinesiologist, measured active wrist ROMs, and active and passive index proximal interphalangeal (PIP) flexion using both goniometers. Wrist flexion/extension and pronation/supination, ulnar/radial deviation, and index finger flexion were evaluated. Before the ROM measurements with goniometers both raters agreed on the anatomic landmarks to be used. In a second occasion (2-5 days later) the ROM measurements were repeated by one of the raters to collect data to test intrarater reliability. Interrater, intrarater, and interinstrument reliabilities were analyzed using statistical tests of differences and Intraclass Correlation Coefficients (ICCs). The construct validity was determined using Pearson's r correlation coefficients from ROM measurements and patient rated pain and function (PRWE) and quick Disabilities of the Arm, Shoulder and Hand (quick DASH) questionnaire. For both electrogoniometers the results revealed high intrarater, interrater and interinstrument reliabilities for most ROM measures (ICC range: 0.64-0.97), and the Bland and Altman plots for mean differences between comparisons did not show any systematic errors. Most of the NK and J-Tech ROM measurements were moderately correlated with the patient rated pain and function scores from DASH and PRWE (r range: 0.32-0.63). There was a significant difference between the torques applied by raters when performing passive ROM measures for the index finger PIP joint. The authors interpreted that this would have a small impact on the overall reliability. The primary purpose of the study included determination of the criterion related validity for the two digital electrogoniometric devices. However, in the limitations section of the article the authors stated that they had "examined reliability and construct validity, but did not measure criterion validity". They also stated that if used a gold standard criterion, such as radiography, their comparisons (between devices, raters) would have had been more accurate. The authors concluded that both goniometers "demonstrated high reliability coefficients and tight error margins". They also concluded that based on the relationship between quick DASH scores and PRWE, 'joint motion impairments' were contributing to 'functional disability'.

Reliability and validity of electrogoniometry measurement of lower extremity movement (Bronner, 2010)[12]

The authors studied reliability, accuracy and validity of flexible electrogoniometers in a sample of 17 advanced dancers (10 female, 7 male). The study was performed in two separate investigations. One was the static test where validity, accuracy, and differences between a digital protractor and electrogoniometers were tested during mid- and end-range movements. The second one was the dynamic test where intrarater and instrument reliability, accuracy, concurrent validity of the electrogoniometers were tested. The 3D Motion analysis was the criterion reference. Intraclass correlations (ICCs) for reliability and standard error of measurement (SEM) for validity were computed and compared to the findings from the 3D Motion analysis. The joints studied were hip, knee, and ankle. The authors employed three biaxial electrogoniometers. Instrument reliability measurements were undertaken on the same day; and intra-rater reliability testing was undertaken in two days. Static test findings (electrogoniometer to protractor correlations) were r ≥0.99, SEM ≤3.65° and dynamic test instrument and intra-rater reliability correlations were r ≥0.98 and r ≥0.97, and SEM ≤3.49° and ≤4.48°, respectively. The concurrent validity correlations (electrogoniometer vs. motion analysis) were high (r ≥0.94). However, as revealed by SEM ≤6.80° (for hip, knee, ankle combined), the dynamic measurement error was also high. Authors concluded that this high level of error "may be acceptable for motion studies or workplace exposure testing where motion analysis is not available or feasible".

Inter-rater, intra-rater and inter-instrument reliability of an electrogoniometer to measure wrist range of motion (da Silva Camassuti, 2015)[14]

The authors studied 24 healthy right-handed volunteers (12 male, 12 female) in their university hand and upper limb clinical research laboratory. Their objective was to test intrarater, interrater and interdevice reliability of a flexible electrogoniometer in measuring wrist range of motion (ROM). They compared the electrogoniometer (MIOTEC) to a universal goniometer (CARCI). To assess interrater reliability two examiners did three series of measurements and to assess the intrarater reliability one of the examiners repeated measurements with electrogoniometer after seven days. They computed Intraclass Correlation Coefficient (ICC), standard error of measurement (SEM) and the Bland-Altman limits of agreement plots for the data analysis. Joint movements studied were extension, flexion, radial and ulnar deviation. Interrater reliability scores determined by ICC were

moderate to excellent (i.e., 0.70-0.79 to higher than 0.90) for both devices. Intrarater reliability was excellent (ICC higher than 0.90) when electrogoniometer and was moderate (ICC 0.70-0.79) when goniometer was used. Interinstrument reliability was moderate when all wrist motions were considered. Bland-Altman limits of agreement was used to check for systematic errors between the mean scores and the difference of scores of all motions. It was poor for interrater assessments using electrogoniometer for ulnar and radial deviation, and for all ROM in terms of interdevice agreements. Nevertheless, the authors concluded that "electrogoniometry is a reliable tool and could be used as an option for clinical applications for wrist motion assessment".

Measurement of Cervical Range of Motion (CROM) by electronic CROM goniometer: A test of reliability and validity (Law, 2013)[23]

Law and Chiu carried out a cross-sectional study to test reliability (intrarater, interrater) and construct validity of an electronic cervical goniometer (composed of dual electronic inclinometers). They recruited 26 patients with cervical pain (who had more than one episode of neck pain in the last three months) and 26 controls (who were free of neck pain in the last six months) from the outpatients of the Physiotherapy Department, the United Christian Hospital, Hong Kong. This convenience sample of 52 people were tested for active ROM (AROM) in three planes; flexion-extension, side flexion, rotation. The measurements were conducted by two physiotherapists with minimum 5-year experience. For interrater reliability, two raters took measurements in two different rooms during the same session. Another measurement session was arranged after seven days to test inrarater reliability. The authors were able to study the construct validity of this electrogoniometer by using a patient group with chronic cervical pain and a control group without pain (Known Group Method). The total cervical AROM in patient group compared to the control group AROM was significantly smaller ($p < 0.001$). This finding reflected the expected difference and supported construct validity. Intraclass Correlation Coefficients (ICC) were computed to test intrarater and interrater reliability for different cervical AROMs in both groups. The results were between 0.75 and 0.92, demonstrating 'good' to 'high' reliability. The authors concluded that "the ACRON electronic cervical goniometer was found to be reliable for measuring cervical mobility in three planes for both normal and patient subjects. Construct validity of the goniometer was supported as the test's result documented significant difference in AROM between the control and the neck pain groups."

The reliability of knee joint position testing using electrogoniometry (Piriyaprasarth, 2008)[24]

The authors conducted two studies to measure inter and intrarater reliability and measurement error when a flexible electrogoniometer is used to assess knee movement. The first study investigated the electrogoniometer during knee movement in sagittal plane at supine, sitting and standing positions (both knees). And the second study, with a detailed protocol, aimed to minimize the measurement error occurring due to reattachment of the electrogoniometer after a 10-meter walk in neutral standing position (right knees only). The study samples included healthy individuals; n=35 (9 male, 26 female; mean age 31) and n=20 (1 male, 19 female; mean age: 20) for the first and second studies, respectively. Measurements were taken by two experienced physiotherapists, of whom one participated in both studies. For study1, intraclass correlation coefficients (ICSs), standard error of measurement (SEMs), and limits of agreement for inter-tester and intratester reliability during knee joint movement measurements were computed across three testing positions (sitting, supine, standing). For study2, SEMs and limits of agreement for inter-tester and intratester reliability of knee joint at neutral standing and after 10-meter walk were computed. The authors found that for the study1 intertester ICCs were 0.58–0.71 (in supine), 0.68–0.79 (in sitting), and 0.57–0.80 (in standing). SEMs between testers was ≤3.55 and limits of agreement were from -12.51° to 12.21°. Intratester ICCs were 0.75–0.76 (in supine), 0.86–0.87(in sitting), and 0.87–0.88 (in standing). SEMs for same tester's repeated measurements were ≤1.7°. Limits of agreement were -8.13° to 7.90°. In study2 intertester reliability based on SEMs ranged from 0.5° to 3.3°, and limits of agreement were -6.0° to 4.9°. Intrarater reliability, measured by SEMs (during standing and after 10-min walking) were 1.3° to 2.3°; and limits of agreement was -6.7° to 2.9°. The authors concluded that using a standard goniometer attachment protocol during measurement of the knee joint movement by a flexible goniometer in standing, supine and sitting positions may help minimize measurement error and may increase reliability.

Estimating the elbow carrying angle with an electrogoniometer: acquisition of data and reliability of measurements (Zampagni, 2008)[25]

The authors conducted an experimental study to test the reliability of elbow carrying angle with an electrogoniometer. They gathered a study sample of 17 men and 20 women (n=37) who were master swimmer athletes in the past and currently had no symptoms related with shoulder, elbow, or wrist

joints. The age range was 41 to 81 years and all subjects gave informed consent for the study. Except for two subjects, all measurements were taken at both arms (right and left) and the total number of measurements was 72. The authors used Faro Arm (electrogoniometer) for measuring the elbow carrying angle. Two orthopedic surgeons, one with more experience than the other, each operated the device. Differences in acquisition by gender or by right and left arms was determined using Student's t-test for independent samples and for paired samples, respectively. Because of great individual differences in carrying angle measurements, they repeated testing with nonparamentric Mann-Whitney and Wilcoxon tests. Neither limb side ($p=0.76$ with t-test and $p=0.74$ with nonparametric test) nor gender ($p=0.57$ with t-test and $p=0.50$ with nonparametric test) differences were significant. Mean carrying angle was 12.7 ± 3.8 ° demonstrating great individual variability with a wide standard deviation. Intraclass Correlation Coefficients (ICC) were computed for intrarater and interrater reliability. Interrater reliability was good ($ICC=0.66$) and intrarater reliability was excellent ($ICC=0.85$). The authors concluded that this method, measuring carrying angle with an electrogoniometer, was fast and suitable for clinical applications.

Summary

- Electrogniometers have been used both in clinical and research settings.
- Wider usage of electrogoniometers is limited with the cost and logistics in employing the device.
- Currently, there are limited number of studies on reliability and validity of electrogoniometers.
- In general, the quality of the reliability/validity studies on electrogoniometers is limited with small size convenience samples, and cross-sectional nature of the study design.
- Often, reliability and validity studies on electrogoniometers are restrained with measurement errors.
- The error may originate from three different sources; technical properties of the electrogoniometer, knowledge/skill of the user (e.g., tester, researcher, clinician), and characteristics of the participant (e.g., patient, study subject).
- Standardizing techniques in ROM measurement may help minimize measurement errors and allow more consistent results even by less experienced testers.
- When reporting on range of motion (ROM) the measurement instrument(s) and procedure(s) should be described in detail. For example, intraclass correlation coefficient (ICC) calculations may differ even for the same tester when working in different study settings.
- As intraclass correlation coefficient (ICC) calculation depends on the population tested, caution is required when decisions on reliability are based on ICCs only. As appropriate, standard error of measurement (SEM) and Bland-Altman limits of agreement should also be provided.
- In general, measurement errors are found to be small when the same rater repeats the measurement using the same device. In the case that measurements are repeated using different instruments and yield similar results, this would be an indication of a valid measurement (e.g., concurrent validity).

Conclusion

Despite limitations with regards to their cost and logistics electrogoniometers are used both in clinical and research settings. Similar to other ROM measurement methods they are also prone to measurement errors. Often, identifying the source of error (e.g., rater, patient, or instrument variations) is difficult. Potential measurement errors can be addressed only if a thorough description of the instruments, assessment methods and analytic approach used is provided. This current literature review was not able to locate high quality studies on reliability and validity of electrogoniometers.

References

1. Demeter, S.L. and G. Andersson, *Disability Evaluation*. 2003: Mosby.
2. Norkin, C.C. and D.J. White, *Measurement of Joint Motion: A Guide to Goniometry*. 2003: F.A. Davis.
3. Radomski, M.V. and C.A.T. Latham, *Occupational Therapy for Physical Dysfunction*. 2014: Wolters Kluwer Health/Lippincott Williams & Wilkins.
4. Chapleau, J., et al., *Validity of goniometric elbow measurements: comparative study with a radiographic method.* Clin Orthop Relat Res, 2011. **469**(11): p. 3134-40.
5. Lavernia, C., et al., *Accuracy of knee range of motion assessment after total knee arthroplasty.* J Arthroplasty, 2008. **23**(6 Suppl 1): p. 85-91.
6. Tousignant-Laflamme, Y., et al., *Reliability and criterion validity of two applications of the iPhone to measure cervical range of motion in healthy participants.* J Neuroeng Rehabil, 2013. **10**(1): p. 69.
7. Bashardoust Tajali, S., et al., *Reliability and Validity of Electro-Goniometric Range of Motion Measurements in Patients with Hand and Wrist Limitations.* The open orthopaedics journal, 2016. **10**: p. 190-205.
8. Santos, C.M.d., et al., *Intra and inter examiner reliability and measurement error of goniometer and digital inclinometer use.* Rev Bras Med Esporte [Internet], 2012. **18**(1): p. 38-41.
9. Werner, B.C., et al., *Validation of an innovative method of shoulder range-of-motion measurement using a smartphone clinometer application.* J Shoulder Elbow Surg, 2014. **23**(11): p. e275-82.
10. Musculoskeletal Key. *Measurement of Range Of Motion and Muscle Length: Background, History, and Basic Principles*. 2016; Available from: https://musculoskeletalkey.com/measurement-of-range-of-motion-and-muscle-length-background-history-and-basic-principles/#bb0510.
11. Foltran, F.A., et al., *Wrist electrogoniometry: are current mathematical correction procedures effective in reducing crosstalk in functional assessment?* Braz J Phys Ther, 2013. **17**(1): p. 32-40.
12. Bronner, S., S. Agraharasamakulam, and S. Ojofeitimi, *Reliability and validity of electrogoniometry measurement of lower extremity movement.* Journal of Medical Engineering & Technology, 2010. **34**(3): p. 232-42.
13. Becari, W. and F.J. Ramirez-Fernandez. *Electrogoniometer sensor with USB connectivity based on the IEEE1451 standard*. in *2016 IEEE International Symposium on Consumer Electronics (ISCE)*. 2016.
14. da Silva Camassuti, P.A., et al., *Inter-rater, intra-rater and inter-instrument reliability of an electrogoniometer to measure wrist range of motion.* Hand Therapy, 2015. **20**(1): p. 3-10.
15. Balogh, I., et al., *Precision of measurements of physical workload during standardized manual handling part III: goniometry of the wrists.* J Electromyogr Kinesiol, 2009. **19**(5): p. 1005-12.
16. Singh, H.P., et al., *Assessment of velocity, range, and smoothness of wrist circumduction using flexible electrogoniometry*. Journal of Hand Surgery, 2012. **37**(11): p. 2331-2339.
17. Armstrong, A.D., et al., *Reliability of range-of-motion measurement in the elbow and forearm.* J Shoulder Elbow Surg, 1998. **7**(6): p. 573-80.
18. Rondinelli, R.D., C.R. Brigham, and E. Genovese, *Guides to the Evaluation of Permanent Impairment - Glossary* 2008, American Medical Association.
19. van Kooij, Y.E., et al., *The reliability and measurement error of protractor-based goniometry of the fingers: A systematic review.* J Hand Ther, 2017. **30**(4): p. 457-467.

20. Sim, J. and P. Arnell, *Measurement validity in physical therapy research.* Phys Ther, 1993. **73**(2): p. 102-10; discussion 110-5.
21. Akizuki, K., et al., *The effect of proficiency level on measurement error of range of motion.* Journal of Physical Therapy Science, 2016. **28**(9): p. 2644-2651.
22. Koo, T.K. and M.Y. Li, *A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.* Journal of Chiropractic Medicine, 2016. **15**(2): p. 155-163.
23. Law, E.Y.H. and T.T.W. Chiu, *Measurement of Cervical Range of Motion (CROM) by electronic CROM goniometer: A test of reliability and validity.* Journal of Back and Musculoskeletal Rehabilitation, 2013. **26**(2): p. 141-148.
24. Piriyaprasarth, P., et al., *The reliability of knee joint position testing using electrogoniometry.* BMC Musculoskeletal Disorders, 2008. **9**(6): p. 22.
25. Zampagni, M.L., et al., *Estimating the elbow carrying angle with an electrogoniometer: acquisition of data and reliability of measurements.* Orthopedics, 2008. **31**(4).

Appendix 1

WorkSafeBC - Evidence-Based Practice Group Levels of Evidence [adapted from 1,2,3,4]

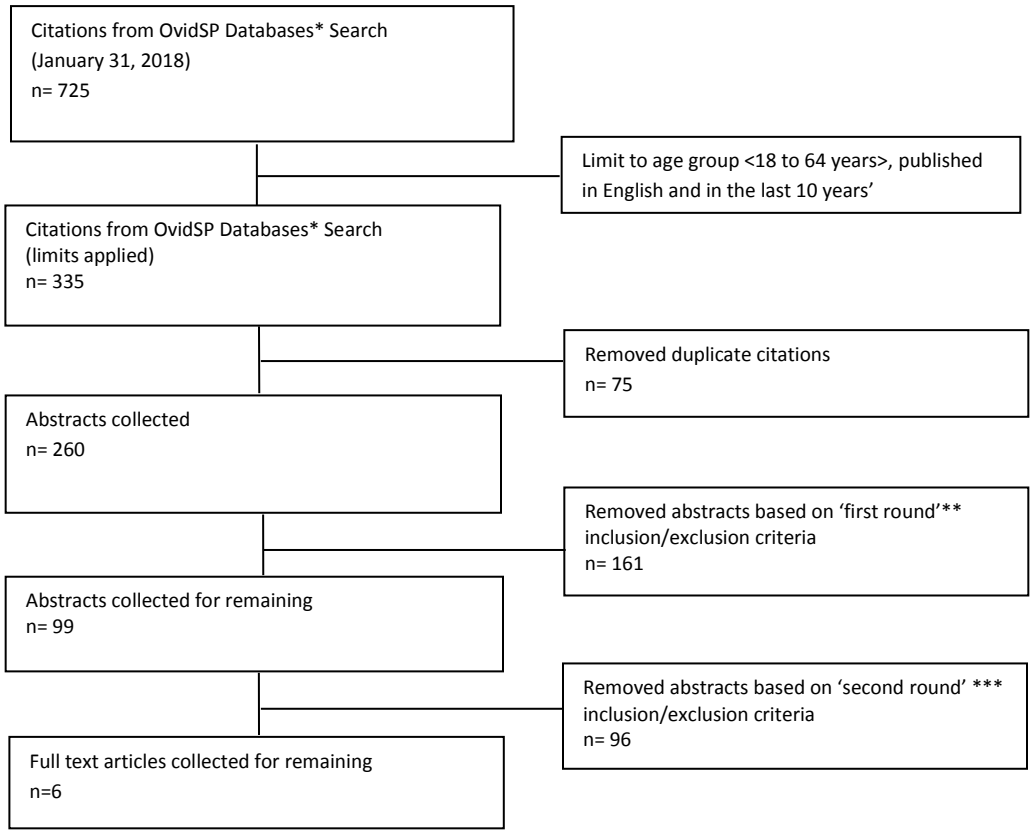
1	Evidence from at least 1 properly randomized controlled trial (RCT) or systematic review of RCTs.
2	Evidence from well-designed controlled trials without randomization or systematic reviews of observational studies.
3	Evidence from well-designed cohort or case-control analytic studies, preferably from more than 1 centre or research group.
4	Evidence from comparisons between times or places with or without the intervention. Dramatic results in uncontrolled experiments could also be included here.
5	Opinions of respected authorities, based on clinical experience, descriptive studies or reports of expert committees.

References

1. Canadian Task Force on the Periodic Health Examination: The periodic health examination. CMAJ. 1979; 121:1193-1254.
2. Houston TP, Elster AB, Davis RM et al. The US Preventive Services Task Force Guide to Clinical Preventive Services, Second Edition. AMA Council on Scientific Affairs. American Journal of Preventive Medicine. May 1998; 14(4):374-376.
3. Scottish Intercollegiate Guidelines Network (2001). SIGN 50: a guideline developers' handbook. SIGN. Edinburgh.
4. Canadian Task Force on Preventive Health Care. New grades for recommendations from the Canadian Task Force on Preventive Health Care. CMAJ. Aug 5, 2003; 169(3):207-208.

Appendix 2

Flow Diagram (Article Selection)



*** OvidSP Databases searched:** Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Health Technology Assessment, NHS Economic Evaluation Database, BIOSIS Previews, Embase, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R)

**** 'first round' inclusion/exclusion criteria**

Include articles on goniometers, on study populations of healthy individuals or patients with musculoskeletal injuries/disorders, in the formats of systematic reviews, meta-analyses, randomized controlled trials, cohort studies, comparative studies, case-control studies, cross-sectional studies, and case series

Exclude articles in the formats of single case reports, conference abstracts, narrative reviews, study protocols, letters to editors; on cadavers, on animal MSDs (musculoskeletal disorders), on pediatric age group MSDs, on stroke patients, on patients with chronic brain and spinal cord conditions (e.g., ankylosing spondylitis, cerebral palsy), on burn patients, on patients using prosthetics, solely on fracture patients, solely on postoperative surgery patients, solely on cost-effectiveness of goniometers

***** 'second round' inclusion/exclusion criteria**

Include articles on validity and reliability of electrogoniometers in ROM assessment

Exclude articles on types of goniometers other than electrogoniometers