Reliability and concurrent validity of knee angle measurement: Smart phone App versus Universal Goniometer used by experienced and novice clinicians.

Milanese Steven (PhD)[1], Gordon Susan (PhD)[1], Buettner Petra (PhD)[1], Flavell Carol (M.Sc)[1], Ruston Sally (PhD)[1], Coe Damien (Bachelor degree (Physiotherapy).) [2], O'Sullivan William (Master degree (Musculoskeletal Physiotherapy))[2], McCormack Steven (Master degree (Clinical Science)) [2]

[1]School of Public Health, Tropical Medicine and Rehabilitation Science

Discipline of Physiotherapy

James Cook University, Townsville, Qld

[2]The Townsville Hospital

Angus Smith Drive

Townsville, Qld

Institution to which the work should be attributed

School of Public Health, Tropical Medicine and Rehabilitation Science

Discipline of Physiotherapy

James Cook University, Townsville, Qld

Corresponding author

Dr. Steven Milanese

School of Public Health, Tropical Medicine and Rehabilitation Science

James Cook University, Townsville, Qld

Telephone: + 61 7 4781 6734. Fax: + 61 7 4781 6868

E.mail: steven.milanese@jcu.edu.au

Keywords: Goniometer, Smartphone, Knee

Introduction

Goniometry is an essential assessment skill in musculoskeletal practice, with the resultant measures used to determine the presence or absence of dysfunction, guide treatment interventions and generate evidence of treatment effectiveness.[1,2]

Universal goniometers (UG) are the most common form of goniometer used in clinical practice.[1,2] They are easily accessible, relatively inexpensive, portable and easy to use.[3] In recent years the advent of smartphones has brought a range of new technological applications (apps) within the reach of most consumers. Smartphones, cellular telephones with built-in applications and internet access,[7] run stand-alone operating system software that provide a platform for application developers.[8] The low cost and user-friendly application interfaces have allowed consumers to access and utilise technologies which were un-imaginable a decade ago. A number of smartphone based goniometry apps are now available[4], with each app utilising a different mechanism for calculating joint angles

With the increased call for accountability of health practitioners to third party funders of health services, and the increasing application of evidence based practice the use of formalised outcome measures has become an important part of clinical practice. Hence, the use of clinically valid and reliable measurement tools to assess joint ROM is an important consideration for therapists.

Reliability studies have shown that on repeated measures the UG demonstrated good overall intra- and inter-tester reliability.[5] Whilst overall reliability of the UG has

been reported as good, the reliability varies according to the joint and the range of movement (ROM) being measured.[6]

The validity of UG measures for knee range of motion have been reported, using measures taken from radiographs as a reference standard.[9] The correlation between universal goniometer measures and radiographs were reportedly higher for larger degrees of knee flexion (*Pearson product-moment correlation coefficient r = 0.73-0.77*) than for smaller degrees of knee flexion (*r = 0.33-0.41*).[9] Whilst used as a reference standard in some studies, issues associated with measuring joint angles from a radiograph, in particular procedural problems associated with the angle of the camera relative to the subject,[1] indicate caution when interpreting and applying these results.

The reliability and validity of UG measures can be affected by incorrect application of the goniometer. Aspects such as the location of bony landmarks, the estimation of the centre of rotation of the joint and ability to locate and maintain the centre of the goniometer over this point, all require attention when using the UG.[1]

In an effort to improve the validity and reliability of the UG, various technologies have been applied to the development of alternative types of goniometers. Studies have examined the use of fluid based goniometers,[10] parallelogram goniometers,[9] biaxial[11] and triaxial,[12] electro-goniometers, computerised goniometers[13] and a digital goniometer.[14] Whilst each form of goniometer has its own inherent benefits, issues such as cost and accessibility mean that the UG remains the equipment of choice for joint angle measurement for most musculoskeletal therapists. Due to the reported

reliability and widespread use, a number of studies have used the UG as the reference standard for validating different goniometers.[9,5]

One available smartphone goniometry app is the Knee Goniometer App (Ockendon©) (KGA). It is an accelerometer based knee goniometer, which measures tibial inclination and then calculates the knee flexion angle using a trigonometric equation. This system differs from other smartphone applications such as the DrGoniometer© which uses a virtual goniometer that is positioned on the smartphone screen on a photograph obtained using the smartphone camera[4].

The KGA requires a one-off calibration against any flat surface. A range of smartphone goniometer apps are available for other joints however the knee was chosen for this study as knee ROM is commonly measured in clinical practice, and has been most commonly used to examine reliability and validity of goniometric tools.

Whilst the KGA designers promote its use to eliminate the difficulties of palpating bony landmarks in the femoral segment, its development was based on certain assumptions; (a) morphologically typical adult patient (b) measurement in the horizontal supine posture and (c) predictable ratio of length femur to tibia (i.e. femur length 1.2 times tibial length).[15] The KGA developers fail to provide a definition of a 'morphologically typical adult patient'. No such assumptions exist for the use of the UG however appropriate anatomical knowledge, eye sight and manual dexterity of the examiner are assumed.

Recent evidence indicates high levels of intra-examiner reliability when measuring maximal knee flexion in healthy participants using the KGA,[16] however no information is available regarding inter-examiner reliability, especially with respect to the clinical experience of the measurer.

The authors of this study observed that the use of smartphone based goniometer apps, such as the KGA, were becoming increasingly popular amongst undergraduate and new graduate physiotherapists. As the results of goniometric measurements are often used to make decisions on clinical management strategies, which may affect the patient's physical, financial, social and psychological well-being, all new instruments designed to measure ROM should be tested thoroughly before use in the clinical setting. Issues such as the intra- and inter-tester reliability of the tool are important as clinical decisions are often based on repeated measures by the same or by different therapists.[8] Errors associated with the use of a goniometer can arise from the tool, the tester or from variability in the performance of the individual.[17]

The purpose of this study was

a) To determine the reproducibility (both intra-tester and inter-tester reliability) of the UG and the KGA for measuring knee ROM.
b) To determine the concurrent validity of the KGA, using the UG as the reference standard.
c) To identify if reliability and concurrent validity values for measurement of knee ROM using a universal goniometer or KGA were altered by the level of experience of the therapists (i.e. observer variability bias).

We hypothesised that there would be agreement between repeated measures of knee ROM when using the UG and the KGA and that the inter-tester and intra-tester reliability of these two instruments would be high.

Method

Ethics approval for the study was obtained from the James Cook University Human Research ethics committee (Ethics approval no: H4062)

Participants

Examiners

Goniometric measurement was performed by three final year students enrolled in the Bachelor of Physiotherapy and three qualified physiotherapists with at least seven years orthopaedic clinical experience, and experience with the use of the UG. None of the students or qualified practitioners had any experience using the KGA. The students had extensive experience with the use of the UG in their undergraduate training.

Subjects to be measured (Measurees)

Measurees for this study were six healthy student volunteers (three men and three women) attending James Cook University, Townsville campus. The right knee was selected for measurement. The measurees were screened by self-report questionnaire, and had no history of musculoskeletal or neurological injury to the lower limb. Each measuree signed an informed consent form prior to participation.

As the aim of the project was to study the reliability of the KGA and UG measurements by different examiners in a normal healthy population there was no attempt to identify if the measurees met the KGA developers assumptions.

With a significance of 0.05 (alpha) and power of 0.20 (beta) and assuming a moderate correlation for UG and KGA measures, we required a sample of 18 joint measures. For pragmatic reasons it was decided to use 18 different jigs across the 6 measurees (i.e. three each) to achieve our 18 joint measures.

Recorders

Documentation of all goniometric measurements was performed by six independent recorders. Recorders were trained in interpretation of the UG angle measuring scale prior to commencement of the data collection.

Instrumentation

Universal goniometers (UG) (Chattanooga©), with plastic 360 degree goniometer face, and 10 inch movable arms were used. One surface of the goniometer face was covered so that the figures could not be seen from the examiners side. The KGA was downloaded onto three Apple iPhone© 4G units. All covers were removed from the iPhones©.

Procedure

The testing session was completed at the Townsville campus of James Cook University. Measurees wore shorts to allow exposure of their right leg, from lateral malleolus to the greater trochanter. They were then placed supine on a standard height adjustable electric treatment plinth. A solid plastic jig was placed under their right knee to ensure a standardised degree of knee flexion between measures. These jigs were triangular in shape and constructed from rigid plastic. Six different jig heights were constructed providing a range of knee flexion angles. Once the measurees had settled onto the jig they were asked not to move for the remainder of the testing session which was approximately 10 minutes. This was monitored by the recorder and no movement of any measuree was observed during testing.

Standard protocols for the use of the UG[18] and KGA[15] were provided to the examiners a week prior to the testing session. On the day of testing all examiners were provided with a familiarisation and training session for both the UG and KGA protocols. When all examiners reported they were confident with the protocol the testing session began.

Universal goniometer protocol (UG)

Examiners were asked to position themselves lateral to the right knee of each measuree. The measurements on the UG were blinded from the examiners at all times. The UG was positioned so that the goniometer axis rested over the lateral epicondyle of the femur. The stationary goniometer arm was aligned parallel to the longitudinal axis of the femur, aligned with the greater trochanter, whilst the mobile arm was placed parallel to the longitudinal axis of the fibula, aligned with the lateral malleolus. When the examiner was satisfied they had completed the measurement, the recorder documented the angle in whole degrees by examining the non-blinded

side of the UG. The recorder ensured no movement of the UG arms occurred during recording.

Knee Goniometer iPhone App protocol (KGA)

Examiners were asked to activate the KGA on the iPhone©. The iPhone© was placed against a true horizontal level (marked on each plinth and checked with a bubble spirit level), and the examiner activated the 'set' facility using the touch screen. The examiners placed the device, with screen facing away, on the subject's lower leg, against the anterior border of the tibia to obtain the measurement. According to manufacturer's instructions the iPhone© may be positioned at any point along the border of the anterior aspect of the tibia. When the examiner was satisfied measurement was complete, they notified the recorder who activated the 'hold' facility using the touch screen. Activation of the 'hold' button stored the goniometric reading for recording purposes. The recorder then documented the measured angle (°s) from the device screen, before clearing the reading from the iPhone©.

Six measurement stations were set up. Each station comprised of a plinth, and one measuree, positioned with their right knee over a jig. Measurees were randomly allocated to a station. One UG or KGA was placed at each of the six stations. The six stations were arranged in a line, with screens between plinths for privacy, and to blind the examiners. The examiners were then allocated to their starting measurement station. The examiner used the allotted UG or KGA, to measure the knee angle and then left the measurement tool at the station before moving to the next measurement station. The examiners moved sequentially through the initial six stations three times. The measurement tools at each station were then altered (i.e.

from UG to KGA or vice versa) and three measurements were undertaken at each station by each examiner using the alternate measurement tool. This resulted in each examiner completing three UG and three KGA measurements at each of the six stations. The complete process was defined as a single 'round' of measures.

All participants then had a 15 minute break following which the measurees, examiners and jigs were randomly allocated to a different plinth and the measurement process was repeated, i.e. another 'round' of measures were completed. This occurred three times in total.

In total each examiner completed three UG and three KGA measurements of 18 different knee positions.

Statistical analysis:

The concordance correlation coefficient (CCC)[19] was used to assess the reliability of UG and KGA measurements within each examiner (three repeat measurements) and between the six examiners. Agreement between UG and KGA measurements was assessed using CCC for each examiner separately and overall. CCCs are presented with 95% confidence intervals (95% CI). The standard error of measurement (SEM) was calculated for all measurements and for each examiner using each tool (UG and KGA).

Scatter and Bland-Altman plots[20] were used to assess agreement visually. The Bland-Altman plot displays a scatter plot of the average UG and KGA measurements versus their differences (UG – KGA measurement). If agreement is good then the

differences should be randomly scattered around the zero difference reference line. Pearson's correlation coefficient was calculated for the averages and differences of the Bland-Altman plot.

A level of significance of 0.05 was assumed. Statistical analysis was conducted using SPSS version 19 (IBM SPSS Inc, Chicago, Illinois).

Results

The mean values (and SD) for experienced clinicians and students are presented in Table 1. The mean, minimum and maximum SEM values were lower for the KGA, although this failed to reach statistical significance. Intra-rater reliability of both UG as well as KGA measurements was high for all examiners, both experienced clinicians and students, with average CCCs all above 0.980 (Table 2). Agreement between UG and KGA measurements was also high for all examiners with average CCCs all above 0.960 (Table 3).

[Insert Table 1 about here]

[Insert Table 2 about here]

[Insert Table 3 about here]

When averaging across repeat measurements as well as examiners, overall agreement was high (CCC = 0.991; 95% CI = 0.979, 0.996) (Figure 1). The Bland-Altman plot showed all but one observation pair to be in the mean +/- 2 standard deviation range of the differences (Figure 2). Pearson's correlation coefficient was -0.51 (p=0.033) suggesting that with increasing measurement values differences between UG and KGA got larger. However, this result was driven by the one

observation pair which had an average difference of -7.39 and when this participant was excluded, correlation was no longer statistically significant (p=0.199).

[Insert Figure 1 about here]

[Insert Figure 2 about here]

Discussion

This is the first study to investigate inter-rater reliability of a new accelerometer based smartphone goniometric application for measuring knee flexion angles.

This study found that both the UG and the KGA displayed excellent reliability over repeated measures of knee joint angles, independent of the level of skill of the operator, i.e. clinician versus final year physiotherapy student, with high overall concordance correlation coefficients (CCC) for averaged measures (three repeat) across all six examiners.

These findings are in agreement with other authors who have reported a high level of intra-rater reliability associated with the universal goniometer[21,6] and maximal knee flexion with the KGA.[16]

Whist the averaged measures showed excellent reliability, when considering the SEM in the repeated measures the range of SEMs across the 18 knee flexion angles were generally larger with the UG than with the KGA, albeit failing to reach statistical significance. This pattern towards an increased range of SEMs with the use of the UG was similar between experienced and novice practitioners. Whilst protocols were developed for both instruments and both groups reported they were

comfortable with their use prior to commencing data collection the need to palpate anatomical landmarks with the UG, particularly the greater trochanter may have resulted in greater potential for error. The placement of the KGA was relatively more stable, being aligned along the bony landmark of the tibia.

From a clinical perspective the SEM in the knee flexion angle measured, using the UG was less than 2.7 degrees, and 1.4 degrees with the KGA. This compares favourably with the variation of 5.5 degrees with the use of the universal goniometer reported by Brosseau et al.[9]

The CCC values for the experienced clinicians were not very different than those of the students, with both groups demonstrating CCCs above 0.96. When considering the differences in measures between the two groups of examiners the mean differences between the experienced clinicians and the students over the 18 measure angles was 2.7 degrees (95% CI 1.1 to 4.1) for the UG and 0.4 degrees (95% CI 0.08 to 1.01) with the KGA. This suggests that the KGA provided more consistent measures between experienced and novice users than the use of the UG. However the degree of difference (under 3 degrees) is less than five degrees which is considered to be the minimal difference which would have a clinical impact.[22]

When considering the use of either goniometric system for measuring a real clinical change in knee flexion angle, improvement has to be considered as greater than the variability reported with repeated measures. The SEM between examiners was up to 2.7^{o} for the UG and 1.4^{o} for the KGA. A determination of real clinical change should be made considering this variability.

When considering the process for validation of the KGA the issue arose about what constituted a suitable reference standard. Whilst radiographic investigations would appear to offer the most rigor in terms of validating a new measuring instrument, the process of ensuring optimal radiographic planar alignment to calculate joint angles has presented a significant hurdle for clinic based research.[1] When considering how goniometers are used in clinical practice their role is to quantify changes in state over time, rather than to identify an absolute value. The criterion for a reference standard is therefore less absolute. To validate a smartphone based goniometric app, such as the KGA, the reference standard selected for this study was a UG, as this reflected the most commonly used form of goniometer in clinical practice. When considering the averaged measure of knee joint flexion angles from the KGA to the UG there were no significant differences, across all six examiners, suggesting that the KGA was a valid mechanism for collecting knee flexion joint angle measures when compared to the UG.

The KGA requires less training, less knowledge of surface anatomy landmarks and less palpation skill. This has advantages for novice practitioners and students and could potentially be used by patients to measure and monitor their own progress. Since the KGA was solely designed to measure knee ROM, this study investigated one joint movement (knee flexion), and no other joints should be examined using this tool.

Many factors have the capacity to alter the effectiveness of joint measurement in the clinical setting for example patient pain, cooperation and anthropometric variation. The use of a jig to standardise the knee position was essential to standardise participant position for repeated measures in this study. The decision to use healthy university volunteers was made for pragmatic reasons, however represents a limitation of this study. The reliability in a clinical setting may be altered by patient pain and cooperation however the results of this study indicate that if patients are able to remain still during measurement the KGA and UG will be reliable. Both the UG and the KGA took similar lengths of time to apply to measure joint angle.

Whilst every effort was made to ensure the subject remained in the same position on the knee angle jigs it was impossible to rule out the potential for some movement to occur between measures. Any changes in joint angle related to this would be expected to be small and non-systematic, and therefore unlikely to significantly alter the direction of any relationship.

This study established that both the UG and the KGA were reliable for measurement of knee flexion angles by experienced clinicians and final year physiotherapy students using standardised protocols. The variability in repeated measures was greater with the UG, however there were no large differences between the measures recorded from both. Small error of measurement values for the UG (< 3 degrees) and the KGA (< 2 degree) might indicate the KGA is superior for assessment where clinical situations demand greater reliability of knee range of motion.

References

1. Gajdosik R, Bohannon R. Clinical measurement of range of motion: Review of Goniometry emphasizing reliability and validity, Physical Therapy. 1987; 67(12):1867-1872.
2. Russell TG, Jull GA, Wooton R. Can internet be used as a medium to evaluate knee angle? Manual Therapy. 2003;8(4):242-246.
3. Croxford P, Jones K, Barker K. Inter-tester comparison between visual estimation and goniometric measurement of ankle dorsiflexion. Physiotherapy Theory and Practice. 1998;14(2):107-113.
4. Ferriero, Giorgio, et al. Reliability of a smartphone-based goniometer for knee joint goniometry. International Journal of Rehabilitation Research. 2013; 36(2):146-151.
5. Brosseau L, Balmer S, Tousignant M, O'Sullivan JP, Goudreault C, Goudreault M, et al. Intra- and intertester reliability and criterion validity of the parallelogram and universal goniometers for measuring maximum active knee flexion and extension of patients with knee restrictions. Archives of Physical Medicine and Rehabilitation.2001; 82:396-402.
6. Rothstein JM, Milleer PJ, Roettger RF. Goniometric reliability in a clinical setting: Elbow and knee measurements. Physical Therapy. 1983;63:1611-1615.
7. PC Magazine [Internet]. Smartphone definition from PC Magazine Encyclopaedia. [cited 2012 Dec 15] . Available from:http://www.pcmag.com/encyclopedia_term/0,2542,t=Smartphone&i=51537,00.asp .

8. Phonescoop [Internet]. Smartphone [cited 2012 Dec 15]. Available from: http://www.phonescoop.com/glossary/term.php?gid=131 .
9. Brosseau L, Tousignant M, Budd J, Chartier N, Duciaume L, Plamondon S, et al. Intra- and intertester reliability and criterion validity of the parallelogram and universal goniometers for active knee flexion in healthy subjects. Physiotherapy Research International. 1997;2:3:150-166.
10. Rheault W, Miller M, Northnagel P, Straessle J, Urban D. Intertester reliability and concurrent validity of fluid based and universal goniometers for active knee flexion. Physical Therapy. 1988;68(11): 1676-1678.
11. Buchholz B, Wellman H. Practical operation of a biaxial goniometer at the wrist joint. Human Factors: The Journal of the Human Factors and Ergonomics Society. 1997;39(1):119-129.
12. Chao E. Justification of triaxial goniometer for the measurement of joint rotation. Journal of Biomechanics. 1980;13(12): 989-1006.
13. Clapper MP, Wolf SL. Comparison of the Reliability of the Orthoranger and the Standard Goniometer for Assessing Active Lower Extremity Range of Motion. Physical Therapy. 1988; 68 (2): 214-218.
14. Carey MA, Laird DE, Murray KA, Stevenson JR. Reliability, validity, and clinical usability of a digital goniometer. Work: A Journal of Prevention, Assessment and Rehabilitation. 2010;36(1): 55-66.
15. Ockendon 2012 [Internet]. Knee Goniometer [cited 2012 Dec 15]. Available from: http://www.ockendon.net/iGoniometer_Home_Page.htm.
16. Hambly K, Sibley R, Ockendon M. Agreement between a novel smartphone application and a long arm goniometer for assessment of

394 knee flexion. International Journal of Physiotherapy and Rehabilitation.
395 2012; 2.
396 17. Piriyaprasarth P, Morris ME. Psychometric properties of measurement
397 tools for quantifying knee position and movement: A systematic review.
398 The Knee. 2004;14:2-8.
399 18. Clarkson H. Musculoskeletal Assessment: Joint Range of Motion and
400 Manual Muscle Strength. 2nd ed. Baltimore: Lippincott Williams &
401 Wilkins;2000.
402 19. Lin IL. A concordance correlation coefficient to evaluate reproducibility.
403 Biometrics. 1989;45(1): 255–268.
404 20. Bland JM, Altman DG. Statistical methods for assessing agreement
405 between two methods of clinical measurement. Lancet. 1986;327(8476):
406 307–310.
407 21. Watkins MA, Riddle DL, Lamb RL, Personius WJ. Reliability of goniometric
408 measurement and visual estimates of knee range of motion obtained in a
409 clinical setting. Physical Therapy. 1991;71:90-96.
410 22. Chaudhary R, Beaupre LA, Johnston DW. Knee range of motion during
411 the first two years after use of posterior cruciate-stabilizing or posterior
412 cruciate-retaining total knee prostheses. A randomized clinical trial.
413 Journal of Bone Joint Surgery American. 2008;90(12): 2579-2586.
414