University of Dayton
eCommons

Physical Therapy Faculty Publications | Department of Physical Therapy

2011

Knee Range of Motion: Reliability and Agreement of 3 Measurement Methods

Paul G. Peters
Wright State University - Main Campus

Michael A. Herbenick
Wright State University - Main Campus

Philip A. Anloague
University of Dayton, panloague1@udayton.edu

Ronald J. Markert
Wright State University - Main Campus

L. Joseph Rubino
Wright State University - Main Campus

Follow this and additional works at: https://ecommons.udayton.edu/dpt_fac_pub

Part of the Biomechanics Commons, Musculoskeletal System Commons, Orthopedics Commons, Sports Sciences Commons, and the Therapeutics Commons

eCommons Citation
Peters, Paul G.; Herbenick, Michael A.; Anloague, Philip A.; Markert, Ronald J.; and Rubino, L. Joseph, "Knee Range of Motion: Reliability and Agreement of 3 Measurement Methods" (2011). *Physical Therapy Faculty Publications*. 32.
https://ecommons.udayton.edu/dpt_fac_pub/32

This Article is brought to you for free and open access by the Department of Physical Therapy at eCommons. It has been accepted for inclusion in Physical Therapy Faculty Publications by an authorized administrator of eCommons. For more information, please contact frice1@udayton.edu, mschlangen1@udayton.edu.

Knee Range of Motion: Reliability and Agreement of 3 Measurement Methods

Paul G. Peters, MS, MD, Michael A. Herbenick, MD, Philip A. Anloague, PT, DHSc, OCS, MTC, Ronald J. Markert, PhD, and L. Joseph Rubino III, MD

Abstract

We conducted a study to compare 3 methods of measuring knee range of motion: visual estimation by physicians, hand goniometry by physical therapists, and radiographic goniometry. We hypothesized that reliability would be high within and across all techniques. We found intrarater and interrater reliability to be satisfactory for visual estimation, hand goniometry, and radiographic goniometry. Interrater reliability across methods did not agree satisfactorily. Between-methods differences in estimating knee range of motion may result from variations in technique among physicians and physical therapists.

Knee range of motion (KROM) is more complex than simple flexion-extension. A significant amount of biomechanical research has been conducted to investigate the intricacies of knee motion.[1-12] Strong static and dynamic stabilizers restrict and coordinate motion of this modified hinge joint. The knee "unlocks" during the initial degrees of flexion, and the femur externally rotates on the tibia. Last[13] indicated that the popliteus creates the external rotation force. In addition, femoral rollback occurs largely through the lateral compartment.[14] However, flexion-extension is a key component; it must be present for normal knee function, with a mean functional arc of 96° and full passive range of motion (ROM) of 135° to 140°.[15,16] Reliable and valid measurement of flexion-extension is important in evaluating surgical outcomes and communicating with therapists.

Dr. Peters is Resident, and Dr. Herbenick is Assistant Professor, Department of Orthopaedic Surgery, Sports Medicine, and Rehabilitation, Boonshoft School of Medicine, Wright State University, Dayton, Ohio.
Dr. Anloague is Program Director, Doctor of Physical Therapy Program, University of Dayton, Dayton, Ohio.
Dr. Markert is Professor, Department of Medicine, and Dr. Rubino is Assistant Professor, Department of Orthopaedic Surgery, Sports Medicine, and Rehabilitation, Boonshoft School of Medicine, Wright State University.

Address correspondence to: Paul G. Peters, MS, MD, Suite 2200, 30 E Apple St, Dayton, OH 45409 (tel, 937-208-2127; fax, 937-208-2920; e-mail, paulpeters44@hotmail.com).

Am J Orthop. 2011;40(12):E249-E252. Copyright Quadrant HealthCom Inc. 2011. All rights reserved.

Numerous investigators have studied static knee joint measurements.[1-4,6,8-12,17] However, opinions continue to vary on the method that should be used to measure KROM. The American Academy of Orthopaedic Surgeons[17] and Rowe[9] supported using visual estimation, while Moore[5] and Salter[10] reported better reliability with hand goniometry. The clinical implications of subtle changes in knee motion have more recently come to bear on this issue. Several anterior cruciate ligament reconstruction studies have linked loss of terminal hyperextension to diminished subjective outcomes and early radiographic arthritic changes.[18,19]

We conducted a study to compare 3 methods of measuring KROM: visual estimation (VE) by physicians, hand goniometry (HG) by physical therapists, and radiographic goniometry (RG). We hypothesized that reliability would be high within and across all techniques.

Methods

Our institutional review board approved this study, and all patients provided informed consent. Twenty-one healthy male volunteers without current knee injury were randomly assigned to have either the left or right knee evaluated. Unilateral knee motion was evaluated to reduce the potential of estimation bias in examining the contralateral knee. Randomization was performed using Research Randomizer v4.0 (http://www.randomizer.org, Geoffrey C. Urbaniak and Scott Plous). All patients underwent physical examination of the knee and documentation of passive supine terminal hyperextension and maximum flexion as measured in degrees with visual assessment. Knee motion examinations were repeated on 13 patients by 3 physicians and 2 physical therapists on the same day, but during a separate session, to allow intrarater reliability evaluation.

Visual estimation was performed with the patient in the supine position and the contralateral knee fully extended. First, for estimation of hyperextension, the examiner placed 1 hand above the knee joint and cupped the contralateral hand behind the heel to lift it off the table until resistance was felt, which was deemed terminal hyperextension (Figure 1). The knee was then flexed by the patient. The examiner stabilized the thigh, and the contralateral hand was placed on the anterior ankle with pressure applied to increase flexion until a firm endpoint was reached and maximum flexion determined (Figure 2). Estimates were recorded to the nearest degree. Three

Figure 1. Visual estimation—hyperextension.

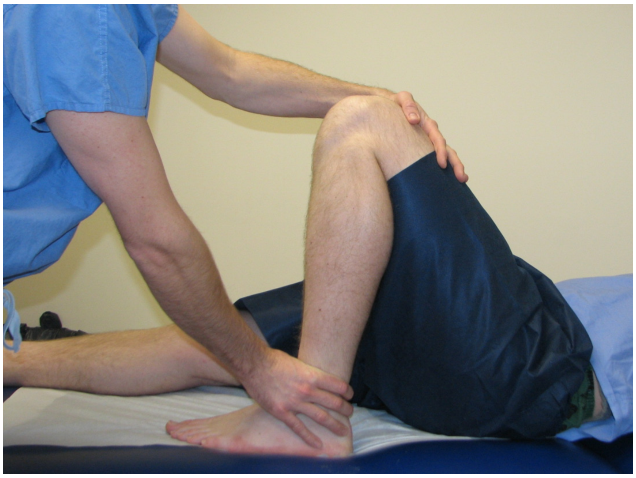

Figure 2. Visual estimation—flexion.

physicians experienced in sports medicine repeated this examination in succession. Data were missing from only 1 physician, who failed to examine 1 patient.

Hand goniometry was performed by 2 experienced physical therapists in succession. Standardized plastic goniometers with 18-cm plastic movable limbs were used with 1° increments for flexion and hyperextension measurements. The patient was supine with the contralateral extremity fully extended. The heel was placed on a small, 12-cm bolster, and goniometric measurement was obtained (Figure 3). The knee was then passively flexed by the therapist into full passive flexion, and the measurement was obtained while the patient held the knee in the position in which it was placed (Figure 4). The physicians and physical therapists were blind to the measurements of their colleagues.

After completion of the physical examination, 3 cross-table lateral radiographs of the knee were obtained. The patient was supine with the contralateral leg in full extension. The hyperextension radiograph consisted of the heel placed on a 12-cm-thick radiolucent bolster with the x-ray plate placed perpendicular to the knee and held in place against the medial skin. The radiologic technician placed the knee into flexion, and the patient held the knee in position by grasping the ankle. The cassette was again placed between the legs against the medial skin of the knee. Finally, a weight-bearing full-squat radiograph was taken. The patient sat on his heels and assumed a baseball catcher's position with the cassette held between the legs. The squat radiograph was added to investigate whether routine physical examination measures true terminal flexion. Whenever a lateral radiograph of the knee proved unacceptable, the radiograph was repeated. Radiographic measurements, made by 2 physicians blind to each other's assessments, involved using a picture archive communication system (Echoes; Med Strat, Downers Grove, Illinois) for electronic medical imaging. Angles were determined using the long axis of the femur and tibia, with approximately 8 cm of shaft visualized, and estimating to the nearest degree through electronic goniometry.

Statistical Analysis

Intraclass correlation coefficient (ICC) was the principal statistic used for this study. ICC can be used to assess both intrarater reliability (agreement within a rater) and interrater reliability (agreement among raters). Satisfactory ICC was set at 0.70. For assessing intrarater reliability of VE and HG, examinations were repeated on 13 patients by 3 physicians and 2 physical therapists. The radiographs for all 21 patients were available for evaluation of agreement for analysis.

We also compared VE, HG, and RG on terminal hyperextension and maximum flexion using the scores from a randomly selected rater for each measurement method. The Shapiro-Wilk test showed that some vari-

Table I. Comparison of 3 Measurement Methods for Flexion and Extension

Measurement, mean (SD)	Visual Estimation (n = 21)	Hand Goniometry (n = 21)	Radiographic Goniometry (n = 21)	*P* Value
Flexion, °	146 (6)	138 (7)	144 (8)	<.001[a,b]
Extension, °	–3.5 (1.7)	–6.3 (3.0)	–4.2 (2.8)	.009[a,c]
Squat flexion, °	—	—	158 (10)	—

[a]Friedman test.
[b]Wilcoxon signed rank test: hand goniometry < visual estimation (P<.001); hand goniometry < radiographic goniometry (P = .004).
[c]Wilcoxon signed rank test: hand goniometry > visual estimation (P<.001); hand goniometry > radiographic goniometry (P = .008).

Figure 3. Hand goniometry—hyperextension.

Figure 4. Hand goniometry—flexion.

ables were not normally distributed. Consequently, we used the nonparametric Friedman test for comparisons of the 3 methods, and the Wilcoxon signed rank test was used for pairwise post hoc comparisons after the Friedman test. A modification of the Bland-Altman approach[20] was used to determine outliers from the mean difference for 2 measurement techniques (HG minus VE, HG minus RG, VE minus RG).

Results

Mean (SD) age of the 21 healthy male patients was 29.6 (4.9) years (range, 22-42 years). Twelve left knees and 9 right knees were examined. During radiographic assessment, 4 studies were repeated because the initial radiograph was unacceptable.

Table I shows that VE, HG, and RG differed on measurement of flexion (P<.001). Although VE (146°) and RG (144°) were nearly identical, HG (138°) was 6° less than RG (P≤.001) and 8° less than VE (P = .004).

Comparison of the 3 methods for measuring extension showed statistical significance as well (P = .009). HG was 2.8° different from VE (P<.001) and 2.1° different from RG (P = .008). Squat flexion was 14° higher than flexion on RG.

Table II lists the ICC results. Intrarater reliability for VE, HG, and RG for extension, flexion, and (RG only) squat flexion was acceptable (all ICCs of intrarater reliability, ≥0.85; median, 0.95). That is, both physicians and physical therapists showed high reproducibility when repeating VE, HG, or RG.

Interrater reliability for RG (2 physicians) was satisfactory for extension (ICC = 0.84), flexion (ICC = 0.99), and squat flexion (ICC = 0.99). For VE (3 physicians), both extension and flexion had an ICC of 0.80. Interrater reliability for HG (2 physical therapists) was 0.88 for flexion but only 0.21 for extension. Interrater reliability across VE, HG, and RG did not agree satisfactorily. Both extension (ICC = 0.45) and flexion (ICC = 0.52) were below the ICC standard of 0.70.

For flexion, HG was lower than VE for all 20 cases and lower than RG in 17 of 21 cases, while VE was lower than RG in 14 of 20 cases. For extension, HG was higher than VE in 11 of 20 cases and higher than RG in 8 of 21 cases, while VE was higher than RG in 5 of 20 cases. Using a modification of the Bland-Altman approach, for flexion, 42 of 61 cases were within 1 SD of the mean difference, and 59 of 61 cases were within 2 SDs. For extension, 41 of 61 cases were within 1 SD of the mean difference, and 59 of 61 cases were within 2 SDs.

Discussion

We found that each rater, regardless of technique used, had high intrarater reliability (ICC > 0.8). Except for HG evaluation of hyperextension, each technique had acceptable reliability across raters. However, there was no agreement across methods. Marks and colleagues[4] were the first to

Table II. Intrarater and Interrater Reliability for 3 Measurement Methods

Reliability[a]	Visual Estimation (VE)[b]		Hand Goniometry (HG)[c]		Radiographic Goniometry (RG)[d]			VE/HG/RG	
	Extension	Flexion	Extension	Flexion	Extension	Flexion	Squat	Extension	Flexion
Intrarater	.95 (.92, .98)	.96 (.93, .99)	.85 (.83, .87)	.97 (.96, .98)	.87 (.75, .99)	.94 (.88, 1.00)	.97 (.96, .98)	—	—
Interrater	.80 (.59, .92)	.80 (.58, .91)	.21 (–.91, .68)	.88 (.70, .95)	.84 (.61, .94)	.99 (.97, .99)	.99 (.98, 1.00)	.45 (–.14, .77).	52 (.00, .79)

[a]95% confidence intervals in parentheses.
[b]Mean intraclass correlation for 3 physicians.
[c]Mean intraclass correlation for 2 physical therapists.
[d]Mean intraclass correlation for 2 physicians.

report adequate agreement between physicians' measurement of KROM using VE in patients with rheumatoid arthritis.

Our results confirmed the high intraobserver and interobserver reliability on KROM assessed through VE. Watkins and colleagues[11] examined the intrarater and interrater reliability of therapists who performed HG and VE of KROM. They found interrater reliability for VE to be 0.83 for flexion and 0.82 for extension compared with 0.90 and 0.86 for flexion and extension, respectively, for HG. They concluded that HG was superior to VE for consistency of measurement. In contrast, our intrarater and interrater reliability for KROM measurements was higher for VE than for HG.

In our study, for flexion, HG was 6° less than RG and 8° less than VE. We later realized that VE flexion was actually estimating forced flexion. This subtle difference in patient positioning resulted in a small but consistent difference that affected our results and that could hinder surgical outcome analysis. Mean radiographic squat flexion was 158°, a mean of 14° more than supine flexion. This difference raises the question as to whether supine flexion is an accurate measurement of maximum flexion. As with the anterior cruciate ligament reconstruction outcome analysis by Shelbourne and Gray,[18] good long-term outcome requires regaining physiologic knee hyperextension. We may be overlooking an important variable by not evaluating true full flexion. Future research on evaluating KROM should address this issue by including squat flexion VE in the physical examination.

Our results should be evaluated with caution for several reasons. First, HG hyperextension had a low interrater reliability estimate. One physical therapist recorded more hyperextension than the second therapist did in 18 of 21 patients. Thus, variation in operator technique affected our ICC results. However, variation in procedure among health care professionals is common in clinical medicine. Our results indicate that subtle differences in estimation are difficult to control. Therefore, standardizing instruction for all individuals who estimate ROM may enhance surgical outcome analysis and communication with physical therapists.

A second reason for caution is that we restricted our sample to younger men to reduce variability and to improve statistical power. Our goal was to evaluate agreement among testing methods. This homogeneous sample reduced the possibility that there would be outliers and other confounding factors that could have affected our results.

Third, assessment of intrarater reliability for VE and HG was done with 13 patients, not all 21 patients, and the second measurement by the physicians and physical therapists took place later the same day. The practical constraints of scheduling prevented the more ideal inclusion of all 21 patients for intrarater reliability assessment with a longer period (perhaps a week or so) between measurements.

Fourth, unilateral KROM was assessed in an attempt to avoid estimation bias. However, contralateral KROM is essential when evaluating surgical outcomes.

Conclusions

We found intrarater and interrater reliability to be satisfactory for 3 methods of KROM evaluation—VE, HG, and RG. Interrater reliability across methods did not agree satisfactorily. Between-methods differences in estimating KROM may result from variations in technique among physicians and physical therapists. Further research is needed to clarify the importance of assessing full squat as possible true full flexion.

Authors' Disclosure Statement

The authors report no actual or potential conflict of interest in relation to this article.

References

1. Brosseau L, Tousignant M, Budd J, et al. Intratester and intertester reliability and criterion validity of the parallelogram and universal goniometers for active knee flexion in healthy subjects. *Physiother Res Int*. 1997;2(3):150-166.
2. Enwemeka CS. Radiographic verification of knee goniometry. *Scand J Rehabil Med*. 1986;18(2):47-49.
3. Gogia PP, Braatz JH, Rose SJ, Norton BJ. Reliability and validity of goniometric measurements at the knee. *Phys Ther*. 1987;67(2):192-195.
4. Marks JS, Palmer MK, Burke MJ, Smith P. Observer variation in examination of knee joints. *Ann Rheum Dis*. 1978;37(4):376-377.
5. Moore ML. The measurement of joint motion; the technic of goniometry. *Phys Ther Rev*. 1949;29(6):256-264.
6. Moore ML. The measurement of joint motion; introductory review of the literature. *Phys Ther Rev*. 1949;29(5):195-205.
7. Piriyaprasarth P, Morris ME. Psychometric properties of measurement tools for quantifying knee joint position and movement: a systematic review. *Knee*. 2007;14(1):2-8.
8. Rothstein JM, Miller PJ, Roettger RF. Goniometric reliability in a clinical setting. Elbow and knee measurements. *Phys Ther*. 1983;63(10):1611-1615.
9. Rowe CR. Joint measurement in disability evaluation. *Clin Orthop*. 1964;(32):43-53.
10. Salter N. Methods of measurement of muscle and joint function. *J Bone Joint Surg Br*. 1955;37(3):474-491.
11. Watkins MA, Riddle DL, Lamb RL, Personius WJ. Reliability of goniometric measurements and visual estimates of knee range of motion obtained in a clinical setting. *Phys Ther*. 1991;71(2):90-96.
12. Wood L, Peat G, Wilkie R, Hay E, Thomas E, Sim J. A study of the noninstrumented physical examination of the knee found high observer variability. *J Clin Epidemiol*. 2006;59(5):512-520.
13. Last RJ. The popliteus muscle and the lateral meniscus. *J Bone Joint Surg Br*. 1950;32:93-99.
14. Freeman MA, Pinskerova V. The movement of the knee studied by magnetic resonance imaging. *Clin Orthop*. 2003;(410):35-43.
15. Hoppenfeld S. *Physical Examination of the Spine and Extremities*. Norwalk, CT: Appleton & Lange; 1992.
16. Scott WN. *Insall & Scott Surgery of the Knee*. 2 vols. 4th ed. Edinburgh, Scotland: Churchill Livingstone; 2006.
17. American Academy of Orthopaedic Surgeons. *Joint Motion: A Method of Measuring and Recording*. Chicago, IL: American Academy of Orthopaedic Surgeons; 1965.
18. Shelbourne KD, Gray T. Minimum 10-year results after anterior cruciate ligament reconstruction: how the loss of normal knee motion compounds other factors related to the development of osteoarthritis after surgery. *Am J Sports Med*. 2009;37(3):471-480.
19. Salmon LJ, Pinczewski LA, Russell VJ, Refshauge K. Revision anterior cruciate ligament reconstruction with hamstring tendon autograft: 5- to 9-year follow-up. *Am J Sports Med*. 2006;34(10):1604-1614.
20. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135-160.

This paper will be judged for the Resident Writer's Award.