Reliability of Ankle Goniometric Measurements

A Literature Review

RobRoy L. Martin, PhD, PT*
Thomas G. McPoil, PhD, PT, ATC†

This article reviews the existing range-of-motion measurement literature related to ankle dorsiflexion and plantarflexion to determine whether the reliability of ankle range-of-motion measurements can be defined, how the characteristics of the study population or clinician affect reliability, and the level of responsiveness for these measures. A MEDLINE search was performed through February 2004, and 11 articles met the inclusion criteria established for this review. Ample evidence was found for intrarater reliability for ankle dorsiflexion and plantarflexion range of motion. Although some evidence for interrater reliability of dorsiflexion was found, little evidence for interrater reliability of plantarflexion range of motion was uncovered. On the basis of the current literature, the responsiveness of ankle joint range-of-motion measurements is uncertain and requires further studies using patient populations. (J Am Podiatr Med Assoc 95(6): 564-572, 2005)

The goniometric measurement of ankle joint range of motion is used in everyday clinical practice to describe impairment. As an evaluative measure, goniometry is used to assess change in status over time that occurs in conjunction with various treatment interventions. For goniometric measures to be interpreted and used in clinical practice for goal setting and documenting a change in status, it is important for the clinician to understand the reliability and responsiveness of these measures.

Reliability is defined as the reproducibility or consistency of a measure. Test–retest reliability is the degree to which the measurements remain stable when repeated two or more times. Intrarater reliability describes the consistency of repeated measurements taken by the same clinician. If intrarater reliability is demonstrated, then one can have confidence in the accuracy of the measurement compared with other measurements by the same clinician under similar conditions. Interrater reliability describes the consistency of repeated measurements taken separately by two or more clinicians. If interrater reliability is demonstrated, then one can have confidence in the accuracy of the measures compared with measurements taken by other clinicians under similar conditions. When interpreting test–retest reliability and determining how this information can be generalized in practice, the amount of time between the measurements, the characteristics of the clinician(s), and the characteristics of the study sample need to be considered. Therefore, test–retest reliability should be considered the degree of consistency of measurement when applied to certain populations under specific measurement conditions.[1]

Percentage of agreement, Cohen κ statistic, Pearson product moment correlation coefficient, and intraclass correlation coefficient (ICC) are statistical procedures that can be used to estimate test–retest reliability. Percentage of agreement is simply the agreement between repeated measures, and the

*Department of Physical Therapy, Duquesne University, Pittsburgh, PA.

†Program in Physical Therapy, Northern Arizona University, Flagstaff.

Corresponding author: RobRoy L. Martin, PhD, PT, Department of Physical Therapy, Duquesne University, 600 Forbes Ave, 111A RSHS, Pittsburgh, PA 15282.

Cohen κ statistic is the level of agreement between repeated measures above chance agreement. Percentage of agreement and the Cohen κ statistic are recommended for nominal- or ordinal-level data.[2] The Pearson product moment correlation coefficient relates to the linear association between measurements and does not take into account systematic bias.[3] The ICC is calculated as the variance between measurements divided by the total variance. The Pearson correlation coefficient and the ICC are recommended for continuous data, such as goniometric measures. Because the ICC relates to the degree to which repeated measurement yields the same score while accounting for systematic bias,[4] it is the most appropriate statistic to estimate the test–retest reliability of goniometric measurements. Intraclass correlation coefficient values range from 0 to 1, with 0 being no agreement and 1 being perfect agreement. An ICC reliability coefficient greater than 0.70 has been considered an acceptable level of reliability for test–retest purposes when applied to clinical measures.[5]

In addition to reliability coefficients, test–retest reliability can be interpreted by the standard error of measurement (SEM). The SEM is defined as follows: $SEM = \sigma(1 - r)^{1/2}$. Using this formula, σ is the SD of the scores and r is the reliability coefficient. The ICC is used to calculate the SEM for determining the expected difference between scores on retesting. The SEM value is defined in the same units as the measurement (in degrees for goniometric measures) and allows for a direct interpretation of reliability. Also, by using the SEM, a confidence interval can be created that allows the clinician to determine the degree of change required to be beyond the measurement error. When interpreting the literature related to reliability, too much emphasis is placed solely on the ICC reliability coefficient. The SEM may be more useful for interpretation of reliability because the calculated ICC may be low while the SEM is still within clinically acceptable limits.

In addition to reliability, responsiveness is also an important consideration for goniometric measures. Responsiveness is the ability to detect clinically important change. The value associated with responsiveness is estimated as the smallest change in a measure that is considered to represent a true change in an individual's condition.[6] Change in condition is defined in one of three ways: observed, estimated, or important.[7, 8] Change that occurs in individuals before and after a treatment or over time that is expected to result in improvement is observed change. For estimated change, a criterion measure of change is used. Change that is important to the patient, clinician, payer, or society defines important change. The change in measurement associated with the change in condition is used to determine a value for a minimal clinically important difference. The study population, type of treatment, timing of data collection, and how the construct for change is defined will affect how the responsiveness of the measure is defined.[8]

The purpose of this project was to review the existing range-of-motion measurement literature related to ankle dorsiflexion and plantarflexion to determine whether 1) the reliability of ankle range-of-motion measurements can be defined, 2) the characteristics of the study population affect reliability, 3) the characteristics of the investigators or clinicians affect reliability, and 4) any evidence of responsiveness for these measures exists.

Methods

A MEDLINE search through February 2004 was performed with combinations of the following key words: *lower leg*, *ankle*, *foot*, *goniometric*, *dorsiflexion*, *plantarflexion*, *evaluation*, *reliability*, and *responsiveness*. Studies were included in the review if they measured ankle dorsiflexion, plantarflexion, or both using a universal goniometer; reported ICC reliability data; involved human subjects; and were published in English. The characteristics of the subjects, the characteristics of the investigators or clinicians, the methods used to measure range of motion, the type of ICC model used for analysis, ICC reliability coefficients, SEM, and responsiveness information were recorded as provided by the authors of the studies. When the original authors did not provide the SEM, it was calculated when possible using the following formula: $SEM = \sigma(1 - r)^{1/2}$.

Results

Eleven articles met the inclusion criteria for this review of the literature.[9-19] Three tables were created to convey the information relating to the methods used in the studies as described by the authors. This information included the following: a description of the subjects, including diagnosis, age, sex, and range of motion; clinician experience; prestudy training; subject position; landmarks used for measurement with goniometer position; and what was assessed in terms of intrarater or interrater reliability, active or passive range of motion, and whether a blinded or standard goniometer was used. The tables were organized on the basis of the characteristics of the subjects who participated in the studies. Tables 1 through 3 provide information related to subjects without pathology, subjects with orthopedic- and diabetic-related di-

Table 1. Methods of the Studies Involving Healthy Subjects

Authors	Subjects	Clinicians	Training	Patient Positioning	Landmarks	Measurement
Kilgour et al[10]	N = 25 Mean age, 10.1 years 11 males and 14 females	1 physical therapist Experience, 30 years 2 other therapists assisted by holding joint position and recording measurement	Pretest training sessions	Supine with knee extended and knee flexed to 90°	Fibula and fifth metatarsal maintain subtalar neutral	Intrarater PROM DF Same day and 7 days apart Blinded goniometer
Van Gheluwe et al[11]	N = 30 Mean age, 24.8 years (20–40 years) 14 males and 16 females	5 podiatric physicians Experience, 1–20 years	No information	Prone with knee extended	Fibula and fifth metatarsal	Interrater PROM DF Same day Standard goniometer
Jonson and Gross[14]	N = 18, from a larger group of 63 Mean age, 21 years (18–30 years) 57 males and 6 females Mean ROM, 16.19° (8–25°)	1 physical therapist Experience, 12 years	Pilot work to standardize methods	Prone with knee extended	Fibula and fifth metatarsal	Intrarater PROM DF 2 days Interrater 2 weeks Standard goniometer
Mecagni et al[12]	N = 8, from a larger group of 34 females Mean age, 74.7 years (64–87 years) Mean ROM: DF, 8.5°; PF, 62.3°	1 physical therapist	No information	Supine with knee extended	Fibula and fifth metatarsal Axis lateral malleolus	Intrarater PROM DF Same day Blinded goniometer
Clapper and Wolf[18]	N = 20 Mean age, 29 years (20.1–40.1 years)	1 physical therapist	No information	Supine with knee extended	No information	Intrarater AROM DF Same day Standard goniometer

Abbreviations: ROM, range of motion; PROM, passive range of motion; AROM, active range of motion; DF, dorsiflexion; PF, plantarflexion.

agnoses, and subjects with neurologic disorders, respectively.

The study by Elveru et al[17] included subjects with neurologic and orthopedic diagnoses, and, therefore, the methodologic descriptions are found in Tables 2 and 3. The 13 subjects with neurologic diagnoses in the study by Elveru et al[17] were reported to have spinal cord injury, head trauma, or cerebrovascular accident. There was no information given on the specific diagnoses for the 37 subjects with orthopedic diagnoses. The methods used by Kilgour et al[10] are described in Tables 1 and 3 because their study included 25 subjects without impairment and 25 with cerebral palsy. The orthopedic-related diagnoses for the subjects in the study by Youdas et al[15] were as follows: ankle ligament sprain (n = 5), knee ligament sprain (n = 15), ankle pain (n = 5), foot pain (n = 10), lateral knee pain (n = 2), patellar pain (n = 7), and anterior leg pain (n = 10).

To assist in defining test–retest reliability, the ICC

Table 2. Methods of Studies Involving Subjects with Diabetes Mellitus and Orthopedic Disorders

Authors	Subjects	Clinicians	Training	Patient Positioning	Landmarks	Measurement
Wilson et al[13]	N = 21 Grade 1 or 2 ankle sprain College athletes 8 males and 13 females	No information	No information	Prone knee flexion	Fibula and fifth metatarsal Axis lateral malleolus	Intrarater PROM, total DF and PF Same day Standard goniometer
Youdas et al[15]	N = 38 General orthopedic disorders Mean age, 26 years (13–71 years) 17 males and 21 females	5 sports medicine and 5 general physical therapists Mean experience, 7 years (5–13 years)	No information	Determined by rater	Determined by rater	Intrarater AROM DF Same day Blinded goniometer
Elveru et al[17]	N = 37 General orthopedic disorders From a larger sample of 50 Mean age, 35.9 years (12–81 years) 22 males and 28 females	14 physical therapists Mean experience, 6.5 years (SD, 3 years)	1 to 2 practice sessions	Determined by rater	Determined by rater	Intrarater Interrater PROM DF and PF Same day Blinded goniometer
Diamond et al[16]	N = 31 Diabetes mellitus Mean age, 59 years (34–77 years) 17 males and 14 females	2 physical therapists	20 training sessions in 18 mo	Prone knee extended	Fibula and fifth metatarsal Axis lateral malleolus Subtalar neutral maintained	Intrarater Interrater PROM DF Same day Standard goniometer

Abbreviations: PROM, passive range of motion; AROM, active range of motion; DF, dorsiflexion; PF, plantarflexion.

reliability coefficients and, if available, SEM values for intrarater dorsiflexion, interrater dorsiflexion, intrarater plantarflexion, and interrater plantarflexion are presented in Tables 4 through 7, respectively. Tables 4 through 7 represent the consistency of measurements taken within 1 to 3 days. Kilgour et al[10] reported that the intrarater reliability values of measures taken 7 days apart were lower than those of measures taken during the same session. The ICC reliability coefficients of measures taken 7 days apart were as follows: 0.7 to 0.75 for subjects without impairment with the knee flexed, 0.75 to 0.9 for subjects with cerebral palsy with the knee flexed, 0.51 to 0.66 for subjects without impairment with the knee extended, and 0.63 to 0.69 for subjects with cerebral palsy with the knee extended. The median absolute differences for measures taken the same day compared with measures taken 7 days apart were significant ($P < .01$). The authors calculated that to be 95% confident that a change had occurred during a 1-week period, range-of-motion values should demonstrate a change of more than 15° for healthy subjects and 20° for subjects with mild spastic diplegic cerebral palsy.[10]

The studies by Kilgour et al[10] and Elveru et al[17] examined the effect of study sample on intrarater reliability. The study by Kilgour et al[10] compared 25 subjects with mild cerebral palsy with age- and sex-matched control subjects without pathology. The results found that the median absolute differences for passive ankle dorsiflexion range-of-motion measures taken on the same day were not significantly different for subjects with cerebral palsy compared with the control group ($P > .05$). Because of these findings, the ICC values for subjects with cerebral palsy and those without impairment were combined. The ICC values for both groups are presented in Table 4.

Table 3. Methods of Studies Involving Subjects with Neurologic Disorders

Authors	Subjects	Clinicians	Training	Patient Positioning	Landmarks	Measurement
Fosang et al[9]	N = 18 Spastic cerebral palsy Mean age, 6.3 years (2–10 years) 7 males and 11 females	6 physical therapists Mean experience, 10.7 years (4–21 years)	1-day training workshop	Supine knee extended	Fibula and fifth metatarsal Subtalar neutral maintained	Interrater PROM DF 3 days apart Rounded measurement to the nearest 5° Standard goniometer
Kilgour et al[10]	N = 25 Spastic diplegic Mean age, 10.1 years 11 males and 14 females	1 physical therapist with 30 years experience took measurements Second therapist held joint position Third therapist recorded the measurements	No information	Supine knee extended and knee flexed	Fibula and fifth metatarsal Subtalar neutral maintained	Intrarater PROM DF Same day and 7 days later Standard goniometer
Elveru et al[17]	N = 13 Neurologic disorders From a larger sample of 50 Mean age, 35.9 years (12–81 years) 22 males and 28 females	14 physical therapists Mean experience, 6.5 years (SD, 3 years)	1 to 2 practice sessions	Determined by rater	Determined by rater	Intrarater Interrater PROM DF and PF Same day Blinded goniometer
Pandya et al[19]	N = 150 Duchenne muscular dystrophy Age range, >1–20 years	5 physical therapists	No information	Supine knee extended	Lateral lower leg and lateral border of foot	Intrarater Interrater PROM DF Same day Rounded measurement to the nearest 5° Standard goniometer

Abbreviations: PROM, passive range of motion; DF, dorsiflexion; PF, plantarflexion.

Elveru et al[17] included subjects with neurologic and orthopedic diagnoses. Because a discrepancy was reported in the reliability between the two groups, the results are presented separately in Tables 4 through 7 (combined, orthopedic, and neurologic groups). Elveru et al[17] found that dorsiflexion measurements had better intrarater and interrater reliability in subjects with neurologic disorders, whereas plantarflexion measurements had better intrarater and interrater reliability in subjects with orthopedic disorders.

The studies by Elveru et al[17] and Van Gheluwe et al[11] examined the effect of clinical experience on reliability. Elveru et al[17] defined an experienced clinician as one who measures ankle dorsiflexion and plantarflexion daily. In the study by Elveru et al,[17] the ICC values for intrarater and interrater reliability of dorsiflexion and plantarflexion differed by less than 0.07 among all clinicians compared with clinicians with more experience. Van Gheluwe et al[11] also noted that no differences in dorsiflexion measurements were

Table 4. Dorsiflexion Intrarater Reliability

Authors	Reliability	SEM
Kilgour et al[10]	ICC[a] = 0.98–0.99 (knee flexed) ICC[a] = 0.96–0.99 (knee extended)	—
Mecagni et al[12]	ICC[a] = 0.97	—
Elveru et al[17,b]	ICC (1,1) = 0.95	—
Diamond et al[16]	ICC (2,1) = 0.93	2.0
Clapper and Wolf[18]	ICC (2,1) = 0.92	—
Youdas et al[15]	ICC (1,1) = 0.64–0.92	0.5–3.9
Pandya et al[19]	ICC[a] = 0.90	—
Elveru et al[17,c]	ICC (1,1) = 0.90	—
Van Gheluwe et al[11]	ICC[a] = 0.86–0.98	0.6–1.7
Fosang et al[9]	ICC[a] = 0.81	4.8
Elveru et al[17,d]	ICC (1,1) = 0.80	—
Jonson and Gross[14]	ICC (2,1) = 0.74	—

Abbreviation: ICC, intraclass correlation coefficient.
[a]Model type not identified.
[b]Subjects with neurologic disorders.
[c]Combining subjects with orthopedic and neurologic disorders.
[d]Subjects with orthopedic disorders.

Table 5. Dorsiflexion Interrater Reliability

Authors	Reliability	SEM
Diamond et al[16]	ICC (2,1) = 0.81	2.5
Elveru et al[17,a]	ICC (1,1) = 0.77	—
Fosang et al[9]	ICC[b] = 0.75	7.3
Pandya et al[19]	ICC[b] = 0.73	—
Jonson and Gross[14]	ICC (2,1) = 0.65	—
Elveru et al[17,c]	ICC (1,1) = 0.50	—
Van Gheluwe et al[11]	ICC[b] = 0.29	3.3
Elveru et al[17,d]	ICC (1,1) = 0	—

Abbreviation: ICC, intraclass correlation coefficient.
[a]Subjects with neurologic disorders.
[b]Model type not identified.
[c]Combining subjects with orthopedic and neurologic disorders.
[d]Subjects with orthopedic disorders.

obtained by clinicians with 7 to 20 years of experience compared with those with 1 year of experience or less. The ICC values for intrarater reliability for passive dorsiflexion rated by experienced and inexperienced clinicians ranged from 0.86 to 0.98 and from 0.87 to 0.95, respectively.

The studies by Kilgour et al,[10] Youdas et al,[15] and Elveru et al[17] examined the influence of measurement technique on reliability. Kilgour et al[10] and Youdas et al[15] compared the reliability of one measurement with the reliability of an average of two measures. Kilgour et al[10] found that averaging two measures of

Table 6. Plantarflexion Intrarater Reliability

Authors	Reliability	SEM
Mecagni et al[12]	ICC[a] = 0.99	—
Clapper and Wolf[18]	ICC (2,1) = 0.96	—
Elveru et al[17,b]	ICC (1,1) = 0.89	—
Elveru et al[17,c]	ICC (1,1) = 0.86	—
Elveru et al[17,d]	ICC (1,1) = 0.72	—
Youdas et al[15]	ICC (1,1) = 0.47–0.96	—

Abbreviation: ICC, intraclass correlation coefficient.
[a]Model type not identified.
[b]Subjects with orthopedic disorders.
[c]Combining subjects with orthopedic and neurologic disorders.
[d]Subjects with neurologic disorders.

Table 7. Plantarflexion Interrater Reliability

Authors	Reliability	SEM
Elveru et al[17,a]	ICC (1,1) = 0.74	—
Elveru et al[17,b]	ICC (1,1) = 0.72	—
Elveru et al[17,c]	ICC (1,1) = 0.6	—

Abbreviation: ICC, intraclass correlation coefficient.
[a]Subjects with orthopedic disorders.
[b]Combining subjects with orthopedic and neurologic disorders.
[c]Subjects with neurologic disorders.

dorsiflexion did not significantly reduce the mean absolute differences for measures taken 7 days apart by one therapist. In contrast, Youdas et al[15] found that taking the average of two measures increased the intrarater reliability for dorsiflexion and plantarflexion. In the study by Youdas et al,[15] the range of ICC values increased from 0.64–0.92 to 0.78–0.96 for dorsiflexion and from 0.47–0.96 to 0.64–0.98 for plantarflexion when the mean of two measures was used for the ICC calculation.

Kilgour et al[10] and Elveru et al[17] also evaluated the effect of subject position on measurement reliability. Kilgour et al[10] did not find a significant difference in same-day intrarater reliability of passive ankle dorsiflexion measures taken when the knee was extended *versus* when the knee was flexed. However, the mean absolute difference that compared same-day measurements with measurements taken 1 week apart (in subjects with cerebral palsy) was significantly lower when the knee was flexed compared with when the knee was extended ($P < .01$). Elveru et al[17] reported that interrater reliability of passive dorsiflexion or plantarflexion was not affected regardless of whether the subject was positioned prone, supine, or sitting and whether the knee was flexed or extend-

ed. When reliability values were calculated separately for clinicians who used the same measurement position *versus* those who used different placements, the type (1,1) ICC values were 0.40 and 0.59, respectively, for dorsiflexion and 0.75 and 0.65, respectively, for plantarflexion.

The study by Wilson et al[13] could not be placed into Tables 4 through 7 because of the measurement techniques used. In their study, total sagittal plane motion was measured, and the loss of range of motion was calculated by subtracting the measurements taken on the injured ankle from those taken on the uninvolved ankle. The type (2,1) ICC calculated for the range-of-motion loss value was 0.88, with an SEM of 4.1°.

Only one study was found to estimate responsiveness. Wilson et al[13] selected a change in loss of range of motion during a 1-week period that started from the third day to the tenth day after ankle sprain as the criterion of change for the estimation of responsiveness. The estimation of responsiveness was calculated as the change in range-of-motion loss divided by $(2 \times SEM^2)^{1/2}$. Thus the minimally reliable change at 95% confidence was calculated to be 11.3°. This implies that a loss of range of motion must be more than 11.3° to distinguish individuals with an acute ankle sprain within the 7-day timeframe. Compared with the minimum observed change of 4.7°, Wilson et al[13] concluded that the loss of range of motion was not reliable enough to detect a change in individuals from the third day to the tenth day after ankle sprain.

Discussion

One of the purposes of this project was to define the reliability of dorsiflexion and plantarflexion goniometric measurements. The preferred method of achieving this would be to combine the results of similar studies in an attempt to summarize the information. However, it was not appropriate to combine the results of any of the studies reviewed because of the differences in the methods used. The differences in study methods also made direct comparison of the investigations difficult. Furthermore, the ability to interpret and apply the results for clinical purposes was problematic for most of the studies reviewed because the authors did not calculate the SEM. Care must also be taken when interpreting and generalizing the results of these studies. Subject characteristics, including age and diagnosis, the amount of training and experience of the evaluators, and the methods used to obtain the measurements must be taken into consideration. For example, information obtained from an elderly diabetic population should not be generalized to a young athletic population. Likewise, high levels of interrater consistency obtained after a training period may not be obtainable by a group of clinicians who have not undergone similar training.

The existing literature provides evidence for intrarater reliability of ankle dorsiflexion and plantarflexion when measured during a short period under various circumstances. Although nine studies (Table 4) provided evidence for intrarater reliability of ankle dorsiflexion measurements, the different ICC models used for the analyses did not allow for direct comparison of the reliability coefficients. If an ICC cutoff value of 0.70 is established as an acceptable level of reliabilty,[5] then all nine studies demonstrate intrarater reliability, irrespective of the subject sample, characteristics of the clinicians, and measurement technique used.

The existing literature also provides evidence for interrater reliability of ankle dorsiflexion. Four of six studies reviewed (Table 5) met the ICC cutoff value of 0.70. Elveru et al[17] reported a type (1,1) ICC of 0 for subjects with general orthopedic disorders, but they failed to offer an explanation for their unusual value. The results reported by Jonson and Gross[14] and Van Gheluwe et al[11] did not meet the established ICC cutoff value based on the mean interrater absolute difference or the calculated SEM, but these measurements may be reliable enough to be clinically useful. Jonson and Gross[14] found that the mean absolute difference between measures taken by two examiners was 2.44°. Van Gheluwe et al[11] found that the interrater reliability of ankle dorsiflexion was poor (0.29), whereas the SEM was within a clinically meaningful range of 3.3°. This result implies that passive ankle dorsiflexion measures taken by different examiners will be within 7° of each other 95% of the time. Three of the four studies that met the 0.70 cutoff value[9, 16, 17] and the study by Jonson and Gross[14] used some type of training session to standardize measurement techniques. Although the study by Kilgour et al[10] had an acceptable ICC value, the SEM values reported for the cerebral palsy subjects were high.

Four studies examined intrarater reliability of ankle plantarflexion. Three[12, 17, 19] of the four studies met the 0.70 cutoff value. In the study by Youdas et al,[15] only three of ten clinicians did not individually meet the 0.70 cutoff value. In that study, each clinician selected their preferred measurement technique. Only the study by Elveru et al[17] examined interrater reliability of ankle plantarflexion. In that study, although the neurologic disorders subject group did not meet the 0.70 cutoff value, the orthopedic disorders group and the combined group (neurologic and orthopedic subjects) were above the cutoff values.

When considering ICC values, a low reliability coefficient may result from a lack of variability within the sample. This can be especially true in studies that use subjects without impairment, as in the study by Van Gheluwe et al.[11] Inconsistencies between clinicians, including misreading the goniometer, using incorrect landmarks, and improperly determining the end range of motion, are potential sources of error that will reduce the reliability coefficient.

In addition to defining the reliability associated with ankle goniometry, the effects on measurement reliability of the subject population, clinician characteristics, and measurement technique were also examined. Subject diagnoses did not seem to appreciably influence the reliability of intrarater measurements, with the exception of individuals with cerebral palsy. In the study by Kilgour et al,[10] although the ICC value was acceptable, the SEM value may have been too high to make the measurement clinically useful for the assessment of dorsiflexion range of motion. A high SEM for interrater reliability of dorsiflexion was also found by Fosang et al[9] in a similar subject population. The high SEM values in subjects with cerebral palsy may be the result of increased muscle tone, leading to measurement variability. Muscle tone should be considered in other neurologic subject populations when ankle joint range of motion is being evaluated. A conclusive determination of how the characteristics of a particular subject population can affect interrater reliability could not be made. With respect to clinical experience and the technique used for measurement, three studies suggested that reliability does not depend on the experience of the clinician performing the measurement, and two studies suggested that reliability will not be affected by measurement technique. The use of training sessions seems to improve the interrater reliability of the measurement of ankle dorsiflexion.

On the basis of the review of existing literature, the responsiveness of ankle dorsiflexion and plantarflexion requires further study under varied clinical scenarios. When interpreting range-of-motion values and changes in range of motion, the most important question to be addressed is how range of motion affects function. The relationship between impairment and functional status should also be questioned. Measures of knee impairment, including laxity, range of motion, and strength, have not been shown to cause limitations in activity.[20, 21] Measures of lumbar-spine impairment are also believed to have little relation to functional ability.[22] However, the relationship between impairment measures of the foot and ankle, specifically ankle range of motion, and functional ability seems to be more substantiated. Laaveg and Ponseti[23] found that ankle range of motion correlated to a functional rating scale in subjects with clubfoot. A variety of studies[24-26] have found that limited ankle movement relates to gait deviations. Mueller et al[27] reported that gait deviations, as a result of limited ankle range of motion, include a slower walking speed and a shorter stride length. A loss of plantarflexion range of motion was compensated for by increasing knee flexion range of motion, whereas limited ankle dorsiflexion range of motion was compensated for by increased external rotation of the lower-extremity limb while walking.[28] Contrary findings have also been reported in studies[13, 29, 30] that found a weak relationship between physical impairment and disability in subjects with grades I and II ankle sprains.

Unlike in experimental studies, a grading system for studies that examine test–retest reliability could not be found. Therefore, it was not possible to formulate a hierarchy to estimate the value of these studies. Each study offers information that may be applied in clinical situations similar to the one described. Clinicians should be cautious when applying the results of studies that used healthy, noninjured subjects because these groups of subjects do not receive any specific treatment.

Conclusion

In this review of the existing literature, ample evidence was found for intrarater reliability for ankle dorsiflexion and plantarflexion range of motion, and there was some evidence for interrater reliability of dorsiflexion as well. Little evidence was found for interrater reliability of plantarflexion range of motion. On the basis of the current literature, the responsiveness of ankle joint range-of-motion measurements is uncertain and requires further studies using patient populations.

References

1. Streiner DL, Norman GR: *Health Measurement Scales: A Practical Guide to Their Development and Use*, 2nd Ed, Oxford Medical Publications, New York, 1995.
2. Huck SW, Comer WH: "Reliability and Validity," in *Reading Statistics and Research*, 2nd Ed, p 84, Harper Collins, New York, 1996.
3. Denegar CR, Ball DW: Assessing reliability and precision of measurement: an introduction to intraclass correlation and standard error of measure. J Sport Rehabil **2:** 35, 1993.
4. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychol Bull **86:** 420, 1979.
5. Lahey MA, Downey RG, Saal FE: Intraclass correlations: there is more than meets the eye. Psychol Bull **93:** 586, 1983.

6. Stratford PW, Binkley FM, Riddle DL: Health status measures: strategies and analytic methods for assessing change scores. Phys Ther **76:** 1109, 1996.
7. Beaton DE, Bombardier C, Katz JN, et al: Looking for important change/differences in studies of responsiveness. J Rheumatol **28:** 400, 2001.
8. Beaton DE: Understanding the relevance of measured change through studies of responsiveness. Spine **25:** 3192, 2000.
9. Fosang AL, Galea MP, McCoy AT, et al: Measures of muscle and joint performance in the lower limb of children with cerebral palsy. Dev Med Child Neurol **45:** 664, 2003.
10. Kilgour G, McNair P, Stott NS: Intrarater reliability of lower limb sagittal range-of-motion measures in children with spastic diplegia. Dev Med Child Neurol **45:** 391, 2003.
11. Van Gheluwe B, Kirby KA, Roosen P, et al: Reliability and accuracy of biomechanical measurements of the lower extremities. JAPMA **92:** 317, 2002.
12. Mecagni C, Smith JP, Roberts KE, et al: Balance and ankle range of motion in community-dwelling women aged 64 to 87: a correlational study. Phys Ther **80:** 1004, 2000.
13. Wilson RW, Gieck JH, Gansneder BM: Reliability and responsiveness of disablement measures following acute ankle sprains among athletes. J Orthop Sports Phys Ther **27:** 348, 1998.
14. Jonson SR, Gross MT: Intraexaminer reliability, interexaminer reliability, and mean values for nine lower extremity skeletal measures in healthy naval midshipmen. J Orthop Sports Phys Ther **25:** 253, 1997.
15. Youdas JW, Bogard CL, Suman VJ: Reliability of goniometric measurements and visual estimates of ankle joint active range of motion obtained in a clinical setting. Arch Phys Med Rehabil **74:** 1113, 1993.
16. Diamond JE, Mueller MJ, Delitto A, et al: Reliability of a diabetic foot evaluation. Phys Ther **69:** 797, 1989.
17. Elveru RA, Rothstein JM, Lamb RL: Goniometric reliability in a clinical setting: subtalar and ankle joint measurements. Phys Ther **68:** 672, 1998.
18. Clapper MP, Wolf SL: Comparison of the reliability of the Orthoranger and the standard goniometer for assessing active lower extremity range of motion. Phys Ther **68:** 214, 1988.
19. Pandya S, Florence JM, King WM, et al: Reliability of goniometric measurements in patients with Duchenne muscular dystrophy. Phys Ther **65:** 1339, 1985.
20. Pantano KJ, Irrgang JJ, Burdett R, et al: A pilot study on the relationship between physical impairment and activity restriction in persons with anterior cruciate ligament reconstruction at long-term follow-up. Knee Surg Sports Traumatol Arthrosc **9:** 369, 2001.
21. Snyder-Mackler L, Fitzgerald GK, Bartolozzi AR, et al: The relationship between passive joint laxity and functional outcome after anterior cruciate ligament injury. Am J Sports Med **25:** 191, 1997.
22. Simmonds MJ, Olson SL, Jones S, et al: Psychometric characteristics and clinical usefulness of physical test in patients with low back pain. Spine **23:** 2412, 1998.
23. Laaveg SJ, Ponseti IV: Long-term results of treatment of congenital club foot. J Bone Joint Surg Am **62:** 23, 1980.
24. Mueller MJ, Minor SD, Sahrmann SA, et al: Differences in the gait characteristics of patients with diabetes and peripheral neuropathy compared with age-matched controls. Phys Ther **74:** 299, 1994.
25. Olney SJ, Griffin MP, McBride ID: Temporal, kinematic, and kinetic variables related to gait speed in subjects with hemiplegia: a regression approach. Phys Ther **74:** 872, 1994.
26. Winter DA, Patla AE, Frank JS, et al: Biomechanical walking pattern changes in the fit and healthy elderly. Phys Ther **70:** 340, 1990.
27. Mueller MJ, Minor SD, Schaaf JA, et al: Relationship of plantar-flexor peak torque and dorsiflexion range of motion to kinetic variables during walking. Phys Ther **75:** 684, 1995.
28. Biescher AD, Brodsky JW, Pollo FE, et al: Functional outcome and gait analysis after triple or double arthrodesis. Foot Ankle Int **20:** 545, 1999.
29. Dettori JR, Basmania CJ: Early ankle mobilization: part II. A one-year follow-up of acute, lateral ankle sprains. Mil Med **159:** 20, 1994.
30. Dettori JR, Basmania CJ: Early ankle mobilization: part I. The immediate effect on acute, lateral ankle sprains. Mil Med **159:** 15, 1994.