

Reliability of Clinical Tests of Foot and Ankle Characteristics in Older People

Hylton B. Menz, PhD*
Anne Tiedemann, BSc, GradDip*
Marcella Mun-San Kwan, BSc*
Mark Dominic Latt, MBBS*
Catherine Sherrington, PhD*
Stephen R. Lord, PhD*

Lower-extremity problems are common in older people; however, the reliability of clinical tools used to assess foot and ankle characteristics has not been rigorously evaluated. This study evaluated the test–retest reliability of a battery of simple clinical tests of foot and ankle characteristics (tactile sensitivity of the first metatarsophalangeal joint, navicular height, foot length and width, hallux valgus severity, an overall foot problem score, ankle flexibility, ankle dorsiflexion strength, and foot pain) in 31 individuals (13 men and 18 women) aged 76 to 87 years recruited from the community. Three examiners performed the tests on two occasions approximately 2 weeks apart. Intraclass correlation coefficients and coefficients of variation were calculated for continuously scored tests, and the kappa statistic (κ) was used to determine the reliability of hallux valgus severity grading. All of the continuously scored tests had acceptable reliability (intraclass correlation coefficients of 0.64 to 0.98; coefficients of variation of 0.6% to 15.0%), as did hallux valgus severity grading ($\kappa = 0.77$; absolute percentage agreement, 84%). These simple clinical tests can now be used with confidence in clinical and research settings to provide reliable and functionally important information regarding foot and ankle characteristics in older people. (J Am Podiatr Med Assoc 93(5): 380-387, 2003)

Foot problems are common in older people and are associated with impaired mobility and limited ability to perform common household tasks.¹⁻³ Recently, it has also been shown that foot impairment is associated with poor balance and that it is a risk factor for falls.^{4, 5} The main limitations of the available literature on foot problems in older individuals are the lack of consensus as to what actually constitutes a “foot problem” and the variety of approaches that have been used to evaluate the presence of foot deformity, range of motion, and hyperkeratotic lesions.⁶

Furthermore, many studies involving clinical assessments of foot problems in older individuals have not evaluated the reliability of their measurement techniques, which raises doubts as to the generalizability of their results to other clinicians and other clinical settings.

Only two studies have directly evaluated the reliability of foot assessments in older people. Cartwright and Henderson⁷ evaluated the interobserver agreement of five chiropodists in determining the presence or absence of 37 different foot conditions in six older individuals. Agreement was found to be satisfactory to very good for most items, except for hallux rigidus and pronation. Van Gisbergen et al⁸ assessed the level of agreement between two chiropodists in determining the presence or absence of impairments

*Prince of Wales Medical Research Institute, University of New South Wales, Randwick, New South Wales, Australia.

Corresponding author: Hylton B. Menz, PhD, Prince of Wales Medical Research Institute, University of New South Wales, High St, Randwick, New South Wales 2031, Australia.

from the World Health Organization's *International Classification of Impairments, Disabilities, and Handicaps* (now the *International Classification of Functioning, Disability and Health*). The results revealed a very high level of agreement between the two chiropodists, indicating that the classification can be used as a reliable tool for documentation of foot problems. However, the clinical usefulness of the classification is somewhat limited, as many common foot conditions are not included.

The results of these two studies indicate that foot-care specialists can diagnose most foot conditions with a high level of reliability. However, both of these studies used a simple checklist approach that does not provide any means of documenting between-subject variation in the severity of foot conditions. Therefore, as part of a large prospective study of risk factors associated with falls, the authors developed a battery of simple clinical tests of foot sensation, size, structure, deformity, range of motion, strength, and pain that they hypothesize may be associated with mobility limitations and risk of falling in older people.⁶ The criteria used when developing this test battery were that the tests must be feasible for older people to undertake, require minimal equipment, and be able to be performed by research staff with no podiatric background. Before applying these tests to a large sample of community-dwelling older people, this study was conducted to determine their feasibility and test-retest reliability.

Methods

Subjects

Thirty-one individuals (13 men and 18 women) aged 76 to 87 years (mean $\pm$ SD age, 80.8 ± 3.1) residing independently in the community were recruited from a database developed from previous research studies conducted by the authors' group. The Human Studies Ethics Committee of the University of New South Wales gave approval for the study, and informed consent was obtained from all of the participants. The prevalence of self-reported major medical conditions and activities of daily living limitations in the study population are given in Table 1.

Three research staff members with no background in podiatric medicine conducted the tests. Before the study, the primary author (H.B.M.) instructed the examiners in how to perform the tests. The examiners then performed each of the measurements and repeated the process approximately 2 weeks later, with the same examiner assessing the same subject on both occasions. With the exception

Table 1. Prevalence of Self-Reported Major Medical Conditions and Activities of Daily Living Limitations

	No. (%) of Subjects (N = 31)
Medical conditions	
Vascular disease	6 (19)
Diabetes	2 (6)
Stroke	3 (10)
Heart disease	8 (26)
High blood pressure	11 (35)
Low blood pressure	2 (6)
Osteoporosis	8 (26)
Osteoarthritis	12 (39)
Rheumatoid arthritis	2 (6)
Painful feet	6 (19)
4 or more medications	13 (42)
Fall in past 12 months	14 (45)
Activities of daily living limitations	
Difficulty with shopping	1 (3)
Difficulty with housework	7 (23)
Difficulty with cooking	1 (3)

of the foot problem score assessment, these examiners examined 14, 13, and 4 subjects each (H.B.M. conducted the foot problem score assessment at test and retest for the initial 5 subjects recruited into the study as part of the examiner training). The examiners were blinded to the subjects' previous assessment scores. In addition to the tests described in the following subsections, subjects also underwent tests of vision, knee extension and flexion strength, reaction time, and balance; however, these data are not reported here.

Tactile Sensitivity of the First Metatarsophalangeal Joint

Tactile sensitivity was measured using a Semmes-Weinstein-type pressure aesthesiometer.⁹ This instrument contains eight nylon filaments of equal length that vary in diameter. The force required to bend each filament is precalibrated and ranges from 0.0045 to 447 g. The filaments were applied to the plantar aspect of the first metatarsophalangeal joint. Subjects were told that the filament would be placed on their foot when the examiner said either "A" or "B." The subject reported back to the examiner whether he or she felt the filament on "A" or on "B." Tactile threshold was determined using a staircase technique that involved presenting suprathreshold filaments and then applying progressively thinner filaments until the subject could no longer feel them. From this point, the examiner applied progressively thicker filaments until the filament that the subject

could detect was determined. The touch threshold was determined from a minimum of three ascending and descending steps. The pressure exerted by this filament was converted from grams to $\log_{10}$ 0.1 mg, yielding a scale of approximately equal-intensity intervals between filaments.

Navicular Height

Navicular height was measured while the subject was fully weightbearing. The navicular tuberosity was palpated and marked with an ink marking pen, and the height of the navicular tuberosity from the ground was measured in millimeters using a steel ruler. Previous studies^{10, 11} indicate that clinical measurement of navicular height has high intratester reliability and is closely correlated with navicular height determined from lateral radiographs.

Foot Length and Width

The subject placed his or her foot on a piece of graph paper, and the examiner traced around the foot with a pen. Foot length was measured from the tip of the longest toe to the posterior aspect of the heel. Foot width was measured from the widest point of the first and fifth metatarsophalangeal joints. Both measurements were recorded in millimeters.

Hallux Valgus Severity

The presence and severity of hallux valgus were determined using the Manchester Scale, developed by Garrow et al.¹² This instrument consists of standardized photographs of feet with four degrees of hallux valgus (none, mild, moderate, and severe). Full-scale versions of the photographs in the original publication were printed and laminated. To determine hallux valgus severity, the examiner placed the photographs alongside the subject's weightbearing foot and selected which photograph most closely resembled the subject's degree of hallux valgus. The severity of hallux valgus for both feet was documented as 0 (absent), 1 (mild), 2 (moderate), or 3 (severe). Garrow et al reported that the interobserver reliability of podiatric physicians grading hallux valgus was very high ($\kappa = 0.86$).

Foot Problem Score

A simple foot problem scoring system was developed to provide an overall measure of foot deformity. This involved observing and documenting the presence of bony prominences, lesser toe deformities, and hyper-

keratotic lesions on both feet. Each abnormal observation was scored 1 point. Lesser toe deformities were scored according to how many joints were affected. For example, a claw toe, in which the proximal and distal interphalangeal joints are fixed and flexed, was scored 2 points, whereas a hammer toe, in which only the proximal interphalangeal joint is affected, was scored 1 point. Any dorsal or plantar hyperkeratotic lesions associated with the deformity scored an additional point for each lesion. Single hyperkeratotic lesions covering multiple joints were scored according to the number of joints affected. For example, a lesion covering the second, third, and fourth metatarsophalangeal joints was scored 3 points. The combination of this score and the hallux valgus severity score for both feet was recorded as the overall foot problem score. In previous studies involving two podiatric physician testers, Menz and Lord found that the absolute intertester agreement for determining the presence or absence of these conditions ranged from 80% to 95%⁴ and that the intraclass correlation coefficient (ICC) for the overall foot problem score was 0.98 (95% confidence interval [CI], 0.95–0.99).⁵

Ankle Flexibility

Ankle flexibility was measured using a modified version of the lunge test.¹³ The lateral malleolus and the head of the fibula were located and marked with an ink pen. Subjects then stood with their right foot placed alongside a vertically aligned clear acrylic plate inscribed with 2° protractor markings and were instructed to take a comfortable step forward with the left leg. In this position, subjects were instructed to bend their knees to squat down as low as possible while keeping the trunk upright and without lifting the right heel from the ground. The position of the fibular head was marked on the clear acrylic plate, and the angle formed between the lateral malleolus and the fibular head was measured (Fig. 1). The test was completed three times, and the average score was documented as the test result.

Ankle Dorsiflexion Strength

Isometric ankle dorsiflexion strength was measured with the subject sitting on a high chair with the foot strapped securely to a plate with a spring gauge attached.¹⁴ The plate pivoted about the axis of the ankle. The subject's hip was kept in 90° of flexion, the knee in 110° of extension, and the ankle in the neutral position (90° to the leg) during this test (Fig. 2). The subject attempted to maximally dorsiflex the foot

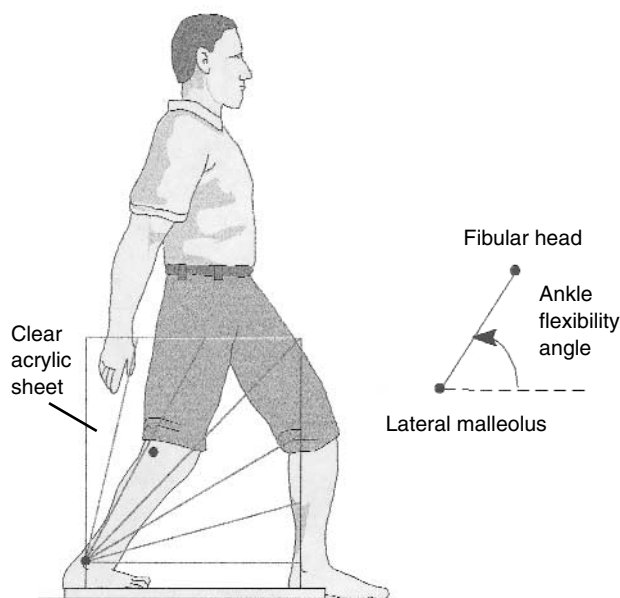


Figure 1. The modified lunge test to assess ankle flexibility.

three times, with intervening rests of 10 to 15 sec, and the greatest force was recorded.

Foot Pain

The Disabling Foot Pain Questionnaire,¹⁵ a 19-item self-completion survey instrument, was used to deter-

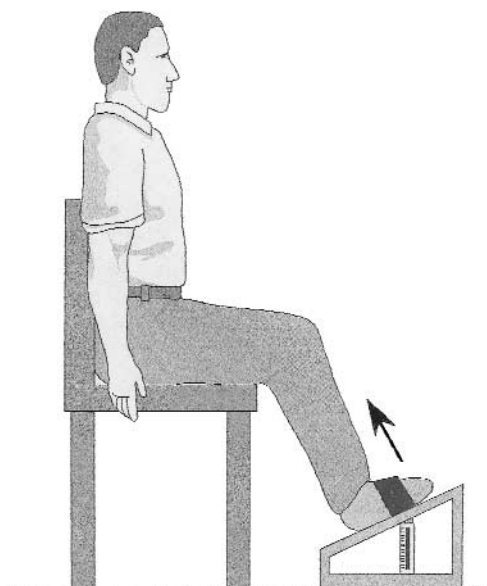


Figure 2. Ankle dorsiflexion strength test.

mine the degree of disability associated with the subject's level of foot pain. The questionnaire has high internal consistency (Cronbach's $\alpha = .99$) and has previously been shown to discriminate between patients attending a rheumatology clinic and a control group of older people. For each of the 19 statements, subjects were asked to indicate whether the statement applied to them "none of the time" (0 points), "on some days" (1 point), or "on most/every day(s)" (2 points), and an overall score was constructed. The resultant questionnaire score ranged from 0 to 38.

Statistical Analysis

Analyses were performed using SPSS version 10 for Windows (SPSS Science, Chicago, Illinois). All data were explored for normal distribution. The foot problem score was positively skewed, so it was $\log_{10}$ transformed before reliability analysis. For the tactile sensitivity, navicular height, foot length and width, foot problem score, ankle flexibility, and ankle dorsiflexion strength variables, ICCs were used to evaluate the test-retest reliability between the two measurement sessions.¹⁶ The type (3,1) ICC model was selected using the two-way mixed-effects model with single-measures options in SPSS.¹⁷ Paired *t*-tests were also performed to test for any systematic differences between sessions, and the coefficient of variation (CV) was calculated to express the variation between trials as a percentage. The test-retest reliability of the hallux valgus severity scale was determined using the kappa (κ) and absolute percentage agreement statistics.

Benchmarks suggested by Fleiss¹⁸ were used to interpret ICC values (>0.75 , excellent reliability; 0.40 to 0.75 , fair to good reliability; and <0.40 , poor reliability). Benchmarks suggested by Landis and Koch¹⁹ were used to interpret the κ ratings (>0.81 , almost perfect; 0.61 to 0.80 , substantial; 0.41 to 0.60 , moderate; 0.21 to 0.40 , fair; 0.00 to 0.20 , slight; and <0.00 , poor).

Results

Reliability

The mean $\pm$ SD time between measurement sessions was 15.9 ± 3.4 days. Paired *t*-tests revealed that there were no systematic differences in mean values for each variable between the two measurement sessions (Table 2).

Test-retest reliability for each of the continuously scored variables is given in Table 3. The ICCs were excellent (0.85 to 0.98) for most variables; however, the ICC for tactile sensitivity and navicular height

Table 2. Scores for the Continuously Scored Variables Across the Two Measurement Sessions

	Score (mean $\pm$ SD [range])	
	Session 1	Session 2
Tactile sensitivity of the first metatarsophalangeal joint ($\log_{10}$ 0.1 mg) ^a	4.6 $\pm$ 0.5 (3.7–5.9)	4.6 $\pm$ 0.6 (3.7–5.0)
Navicular height (mm)	41.7 $\pm$ 6.5 (28–54)	41.7 $\pm$ 6.3 (30–55)
Foot length (mm)	254.2 $\pm$ 17.8 (225–290)	253.5 $\pm$ 17.2 (230–288)
Foot width (mm)	98.4 $\pm$ 9.2 (83–115)	99.3 $\pm$ 9.8 (83–122)
Foot problem score ^b	7.2 $\pm$ 4.8 (1–23)	7.2 $\pm$ 5.3 (1–27)
Ankle flexibility ($^{\circ}$) ^c	38.4 $\pm$ 6.2 (23–55)	37.9 $\pm$ 6.7 (29–57)
Ankle dorsiflexion strength (kg)	11.8 $\pm$ 4.7 (3–20)	12.6 $\pm$ 5.1 (6–27)

^a Pressure exerted by monofilament.

^b Cumulative score for both feet.

^c Measured while fully weightbearing.

measurement were fair to good (0.70 and 0.64, respectively). Coefficients of variation were small for all of the variables (range, 0.6% to 15.0%), indicating that the measurements could be expected to vary little from one session to the next.

The kappa statistic for test–retest reliability of hallux severity grading indicated substantial agreement ($\kappa = 0.77$), with an absolute percentage agreement of 84%. This indicates that the examiners recorded identical scores across the two measurement sessions in all but five subjects. Of these five subjects, there was disagreement between no hallux valgus and mild hallux valgus in four subjects and between mild and moderate hallux valgus in one subject. There were no cases in which severe hallux valgus was scored inconsistently.

Reliability coefficients could not be calculated for the Disabling Foot Pain Questionnaire, as few subjects registered a score (ten in the first session and seven in the second session). Only one subject demon-

strated a significant degree of disabling foot pain, as evidenced by a score of 22 in the first session and 18 in the second session.

Associations Between Tests

Navicular height was significantly associated with foot length ($r = 0.57$; $P < .01$) and foot width ($r = 0.54$; $P < .01$), indicating that if comparisons are to be made between subjects, navicular height will need to be corrected for foot size. There were no other significant associations between test scores.

Discussion

Despite the very high prevalence of foot problems in older individuals, and the proposed relationship among foot problems, mobility impairment, and falls,⁴⁶ few tools have been developed for the assessment of foot problems in the elderly population. The results reported here show that, with brief training, a range of simple clinical tests of foot sensation, size, structure, deformity, range of motion, and strength can be reliably measured by examiners with no background in podiatric medicine and can therefore be used with confidence in research studies and in clinical practice settings.

Before discussing the results of the individual tests, it is worth considering the statistical issues associated with reliability assessment. Although ICCs are generally considered to be more appropriate indicators of reliability than simple correlation coefficients (ie, Pearson's r and Spearman's ρ), a low ICC does not necessarily mean that a test is unreliable. In cases in which there is little variability among subjects' scores, it is difficult to obtain a high ICC, although the scores may not vary considerably between

Table 3. Test–Retest Reliability Results for Continuously Scored Variables

	ICC (95% CI)	CV (%)
Tactile sensitivity of the first metatarsophalangeal joint	0.70 (0.45–0.84)	5.1
Navicular height	0.64 (0.38–0.81)	6.1
Foot length	0.98 (0.95–0.99)	0.6
Foot width	0.85 (0.72–0.93)	2.9
Foot problem score ^a	0.92 (0.84–0.96)	15.0
Ankle flexibility	0.87 (0.74–0.94)	3.6
Ankle dorsiflexion strength	0.88 (0.76–0.94)	9.7

Abbreviations: ICC, intraclass correlation coefficient; CI, confidence interval; CV, coefficient of variation.

^a $\log_{10}$ transformed before reliability analysis.

test and retest. To overcome this, some authors recommend the complementary use of an absolute (as opposed to a relative) measure of reliability such as the CV. The CV expresses the percentage variation of subjects' scores between trials, thereby providing a clinically useful indicator of agreement that is not affected by the presence of a homogeneous sample. Thus a clinical measurement that has only a moderate ICC may still be reliable, provided that the CV is low and there is no systematic difference from one trial to the next.²⁰

The reliability of the measurements obtained in this study is generally in agreement with that of previous studies conducted on younger subjects and different clinical groups. The high reliability of tactile sensitivity measurement of the first metatarsophalangeal joint is consistent with the findings of Diamond et al,²¹ who reported intratester κ values between 0.72 and 0.83 for the assessment of sensory loss at nine sites on the plantar surface of the foot in subjects with diabetes mellitus. Normal aging is associated with peripheral sensory loss,²² and sensory loss is associated with impaired balance and reduced walking speed²³; therefore, the measurement of tactile sensitivity seems to be of considerable importance in determining mobility limitations and risk of falling in older individuals.

Although exhibiting a lower ICC (0.64) than the other measures (owing to the restricted range of values in the sample), navicular height exhibited a low CV (6.1%) and may therefore be a useful measure of foot posture in older individuals. This result is similar to that of Evans et al,²⁴ who reported good reliability of navicular height measurement in adolescents and adults, but is in contrast to that of Vinicombe et al,²⁵ who reported low-to-moderate reliability of navicular displacement measurement (navicular drop and drift) in young adults. The lower reliability in the study by Vinicombe et al is likely due to the variability introduced by measuring subjects in single-limb stance and the fact that two measurements were required to indicate the displacement of the navicular rather than its static position. Whether navicular height when standing is a valid indicator of foot function is yet to be adequately determined^{26, 27}; however, it seems that navicular height can be measured with acceptable reliability and that it is a valid indicator of the height of the arch.^{10, 11}

The high reliability of the lunge test of ankle dorsiflexion is consistent with findings from the study by Bennell et al,¹³ who reported excellent reliability of the lunge test of ankle flexibility in young subjects (ICC, 0.97). However, the lunge test used in the present study differed slightly from that reported by Ben-

nell et al. In that study, subjects were instructed to lunge forward against a wall multiple times until the maximum distance from the wall where they could touch the wall with their knee while maintaining heel contact was determined. As this was considered to be too fatiguing and potentially destabilizing for elderly subjects, the current authors used an alternative protocol in which subjects took a comfortable step forward and squatted down as far as possible while maintaining heel contact. Although subjects were allowed to lean on a bench for support, this did not seem to affect the degree to which they could squat down. Furthermore, ankle flexibility scores were not associated with knee extension strength or balance scores, indicating that performance on this test is not limited by the strength of the quadriceps muscles and is not affected by the subject's balance ability. Given the protocol used in the present study, the values reported here cannot be directly compared with those of Bennell et al or with other techniques of measuring range of ankle dorsiflexion,²⁸ particularly those that test subjects with the knee extended (thereby assessing the contribution of gastrocnemius tightness). However, the authors feel that the technique provides a useful indicator of overall ankle flexibility that is feasible for older people to undertake.

The Manchester Scale¹² for grading hallux valgus severity was found to be easy to use and reliable, as indicated by a κ of 0.77 and an overall absolute percentage agreement of 84%. Although there was some error involved in grading absent *versus* mild hallux valgus and mild *versus* moderate hallux valgus, the examiners documented severe hallux valgus with 100% consistency. The use of life-sized reproductions of the photographs provided in the original publication allowed the examiners to perform a simple visual matching task instead of direct angular measurement of the subject's foot, thereby avoiding the observer bias associated with goniometric assessments. It is likely that the scale provides a reasonably accurate estimation of the hallux abductus angle; however, a comparison of the severity rankings with measurements obtained from radiographs would be required to confirm the validity of this assumption.

The present authors could not adequately determine the reliability of the Disabling Foot Pain Questionnaire,¹⁵ as few of the subjects experienced foot pain. In addition, during the study some problems were experienced with the instrument. Although subjects were advised that the questionnaire related specifically to foot pain, several people attempted to respond to the statements from the perspective of their general health, necessitating clarification by the ex-

aminer and subsequent correction of the responses. The authors of the Disabling Foot Pain Questionnaire acknowledged this problem during the initial development of the instrument, and to emphasize that the questions related to foot pain, the phrase "Because of pain in my feet" was highlighted in the final version. However, this strategy does not seem to be completely effective. Future applications of the questionnaire may need to add the stem "Because of pain in my feet" to the beginning of every statement to remind subjects of the specific nature of the questions. In addition, the test-retest reliability of the questionnaire will need to be evaluated in a population with a higher prevalence of foot pain than the group described here.

The study findings need to be viewed in light of several limitations. First, this investigation focused on intratester reliability, so the level of agreement between different examiners is yet to be addressed. Generally, intertester reliability of clinical measurement is lower than intratester reliability owing to variations in the application and interpretation of the tests between examiners. Although the tests reported here are reasonably objective, some require clinical examination skills that may be applied in a variable manner. For example, differences in manual palpation may influence the reliability of navicular measurement,²³ and differences in the degree of encouragement provided to the subject may affect the measurement of ankle dorsiflexion strength. These issues will need to be addressed if the tests are to be performed by different examiners. The other limitation is that although the examiners had no podiatric qualifications, all had considerable experience in conducting tests of physical function in older people, so the results reported here may not be generalizable to examiners without similar experience.

Conclusion

The results of this study show that if simple, objective, and clearly defined tests are used and staff members are appropriately instructed in the use of the tests by a podiatric physician, examiners without a background in podiatric medicine can reliably measure foot and ankle characteristics in older people. This is of considerable practical benefit in the context of large-scale epidemiologic studies, as it indicates that research staff need not have podiatric qualifications to obtain reliable assessments. These tests can now be used with confidence in clinical and research settings to provide functionally important information regarding foot and ankle characteristics in older individuals.

Acknowledgment. This project was funded by Health Research Partnerships grant 209799 (Prevention of Injuries in Older People) from the National Health and Medical Research Council. Dr. Menz is currently a National Health and Medical Research Council Australian Clinical Research Fellow.

References

1. BLACK JR, HALE WE: Prevalence of foot complaints in the elderly. *JAPMA* **77**: 308, 1987.
2. BENVENUTI F, FERRUCCI L, GURALNIK JM, ET AL: Foot pain and disability in older persons: an epidemiologic survey. *J Am Geriatr Soc* **43**: 479, 1995.
3. GORTER KJ, KUYVENHOVEN MM, DE MELKER RA: Nontraumatic foot complaints in older people: a population-based survey of risk factors, mobility, and well-being. *JAPMA* **90**: 397, 2000.
4. MENZ HB, LORD SR: Foot pain impairs balance and functional ability in community-dwelling older people. *JAPMA* **91**: 222, 2001.
5. MENZ HB, LORD SR: The contribution of foot problems to mobility impairment and falls in community-dwelling older people. *J Am Geriatr Soc* **49**: 1651, 2001.
6. MENZ HB, LORD SR: Foot problems, functional impairment, and falls in older people. *JAPMA* **89**: 458, 1999.
7. CARTWRIGHT A, HENDERSON G: *More Trouble with Feet: A Survey of the Foot Problems and Chiropody Needs of the Elderly*, Department of Health and Social Security, HMSO, London, 1986.
8. VAN GISBERGEN MJ, DEKKER J, ZUIJDERDUIN W: Reliability of the diagnosis of impairments in survey research in the field of chiropody. *Disabil Rehabil* **15**: 76, 1993.
9. SEMMES J, WEINSTEIN S, GHENT L, ET AL: *Somatosensory Changes After Penetrating Brain Wounds in Man*, Harvard University Press, Cambridge, MA, 1960.
10. SALTZMAN CL, NAWOCZENSKI DA, TALBOT KD: Measurement of the medial longitudinal arch. *Arch Phys Med Rehabil* **76**: 45, 1995.
11. WILLIAMS DS, MCCLAY IS: Measurements used to characterize the foot and the medial longitudinal arch: reliability and validity. *Phys Ther* **80**: 864, 2000.
12. GARROW AP, PAPAGEORGIOU A, SILMAN AJ, ET AL: The grading of hallux valgus: the Manchester Scale. *JAPMA* **91**: 74, 2001.
13. BENNELL KL, TALBOT RC, WAJSWELNER H, ET AL: Intratester and inter-rater reliability of a weightbearing lunge measure of ankle dorsiflexion. *Aust J Physiother* **44**: 175, 1998.
14. LORD SR, CLARK RD, WEBSTER IW: Postural stability and associated physiological factors in a population of aged persons. *J Gerontol* **46**: M69, 1991.
15. GARROW AP, PAPAGEORGIOU AC, SILMAN AJ, ET AL: Development and validation of a questionnaire to assess disabling foot pain. *Pain* **85**: 107, 2000.
16. SHROUT PE, FLEISS JL: Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* **86**: 420, 1979.
17. NICHOLS DP: Choosing an intraclass correlation coefficient. *SPSS Keywords* No. 67, 1998. Available at: <http://www.spss.com/tech/stat/articles/whichicc.htm>. Accessed July 16, 2002.
18. FLEISS J: *The Design and Analysis of Clinical Experi-*

ments, John Wiley & Sons, New York, 1986.

19. LANDIS J, KOCH G: The measurement of observer agreement for categorical data. *Biometrics* **33**: 159, 1977.
20. PORTNEY LG, WATKINS MP: *Foundations of Clinical Research: Applications to Practice*, Appleton & Lange, East Norwalk, CT, 1993.
21. DIAMOND JE, MUELLER MJ, DELITTO A, ET AL: Reliability of a diabetic foot evaluation. *Phys Ther* **69**: 797, 1989.
22. LORD SR, WARD JA: Age-associated differences in sensori-motor function and balance in community dwelling women. *Age Ageing* **23**: 452, 1994.
23. LORD SR, LLOYD DG, LI SK: Sensori-motor function, gait patterns and falls in community-dwelling women. *Age Ageing* **25**: 292, 1996.
24. EVANS A, COPPER A, SCHARFBILLIG R, ET AL: Reliability of the foot posture index and traditional measures of foot position. *JAPMA* **93**: 203, 2003.
25. VINICOMBE A, RASPOVIC A, MENZ HB: Reliability of navicular displacement measurement as a clinical indicator of foot posture. *JAPMA* **91**: 262, 2001.
26. MENZ HB: Alternative techniques for the clinical assessment of foot pronation. *JAPMA* **88**: 119, 1998.
27. RAZEGHI M, BATT ME: Foot type classification: a critical review of current methods. *Gait Posture* **15**: 282, 2002.
28. ROME K: Ankle joint dorsiflexion measurement studies: a review of the literature. *JAPMA* **86**: 205, 1996.