OsteoArthritis and Cartilage (2007) **15,** 1008–1012

doi:10.1016/j.joca.2007.02.016

Osteoarthritis and Cartilage

International Cartilage Repair Society

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Validation of a self-report instrument for assessment of hallux valgus

E. Roddy B.M.B.S., M.R.C.P., Specialist Registrar in Rheumatology*, W. Zhang Ph.D., Associate Professor of Musculoskeletal Epidemiology
and M. Doherty M.D., F.R.C.P., Professor of Rheumatology
Academic Rheumatology, University of Nottingham, UK

Summary

Objectives: To validate an instrument for self-reported hallux valgus (HV).

Methods: The self-report instrument consists of five line drawings for each foot depicting a sequential increase in HV angle of 15° developed from a photograph of a normal foot. Participants were asked to select the picture which best represented their left and right feet in turn. Four hundred and fifty-nine subjects completed the self-report instrument: 100 attending a hospital rheumatology clinic and 359 who participated in a community questionnaire study. Three hundred and eighty-four completed it on two occasions (1–2 months apart in 71 subjects and 3–6 months apart in 313) and were assessed once by a blinded observer. Twenty-five subjects were assessed by the blinded observer on two occasions. Validity of the instrument was assessed by the weighted kappa statistic for subject–observer agreement and reliability by the weighted kappa statistics for subject repeatability and observer repeatability. These analyses were repeated for HV dichotomised as present or absent.

Results: For the five-grade HV scale, weighted kappa scores (left and right feet combined) were 0.45 for subject–observer agreement, 0.53 at 1–2 months and 0.51 at 3–6 months for subject repeatability, and 0.82 for observer repeatability. For the dichotomised scale (left and right feet combined), sensitivity was 75% and specificity was 82%: kappa scores were 0.55 for subject–observer agreement, 0.63 at 1–2 months and 0.61 at 3–6 months for subject repeatability and 0.83 for observer repeatability.

Conclusions: The HV self-report instrument provides a valid and reliable assessment of the presence and severity of HV and appears suitable for use in epidemiological studies.

Key words: Hallux valgus, Validation, Self-report.

Introduction

Hallux valgus (HV) is a common deformity characterised by abnormal angulation, rotation and lateral deviation of the great toe at the first metatarsophalangeal (MTP) joint. Estimates of the community prevalence of HV vary widely, ranging from 21% to 70%[1–5]. HV has been reported to be associated with foot pain[1], although other studies have failed to confirm this association[5–7]. It also associates with poor balance[8], immobility[8] and risk of falling[8,9].

An important reason for the inconsistent findings in studies of HV is the widely varying definitions of HV that have been employed, including "bunions", either self-reported or confirmed on physical examination, and observer-assessed HV with no clear description of the definition employed. The lack of a validated assessment tool for self-reported HV is a barrier to large-scale epidemiological studies of HV. Although a method of grading HV by comparison to a series of photographs has been shown recently to be a reliable measure when used by research staff and podiatrists[10,11] and to correlate with radiographic HV[12], it has not been validated for self-reported HV.

*Address correspondence and reprint requests to: Edward Roddy, Primary Care Musculoskeletal Research Centre, Primary Care Sciences, Keele University, Staffordshire ST5 5BG, UK. Tel: 44-1782-583905; Fax: 44-1782-583911; E-mail: e.roddy@cphc.keele.ac.uk

Received 25 August 2006; revision accepted 10 February 2007.

This study was undertaken to develop and validate a line-drawing instrument for the assessment of self-reported HV.

Methods

The study was approved by the Nottingham Local Research Ethics Committees 1 and 2. According to the requirements of the Research Ethics Committees, verbal consent to participate was obtained from subjects participating in the hospital pilot and written consent from those recruited from the community.

DEVELOPMENT OF THE SELF-REPORT INSTRUMENT

A self-report instrument consisting of line drawings depicting varying degrees of HV was developed from a photograph of a normal foot without HV (Fig. 1). Each of the five drawings illustrated a sequential increase in HV angle of 15°. The drawings were accompanied by instructions for subjects to compare the line drawings to their own barefeet without shoes and socks whilst standing and to select the picture which best represents their left and right feet in turn.

VALIDATION OF THE SELF-REPORT INSTRUMENT

The instrument was validated in two groups of subjects. An initial pilot was undertaken in patients attending the

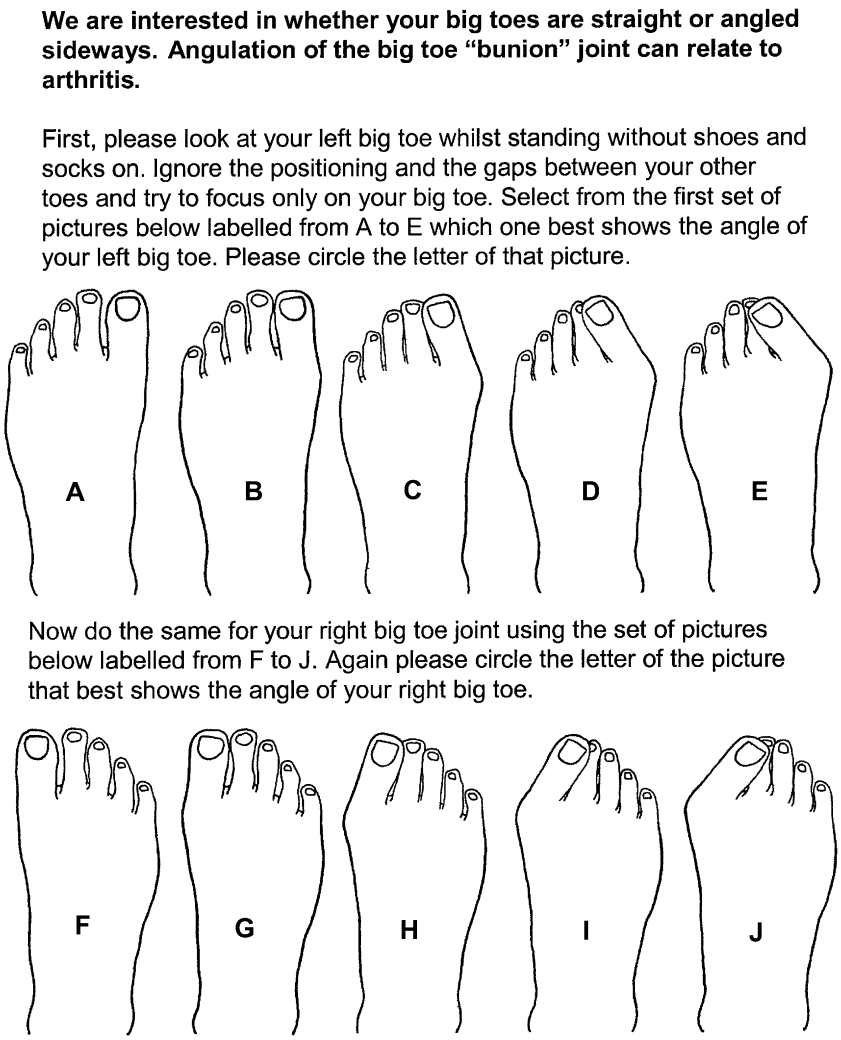

Fig. 1. Self-report instrument for assessment of HV.

rheumatology clinic at Nottingham City Hospital over a 5-day period. All patients attending the clinic, both new and follow-up, were provided with a study information sheet and then invited to participate by their clinician. The first 100 subjects who agreed to participate were enrolled. The subjects completed the self-report instrument and a blinded observer assessed the degree of HV from as close to the subject's perspective as possible using the same line drawings. Twenty-five subjects were seen again 3–6 months later when both subject and observer assessments were repeated.

The study was then extended to 359 patients who participated in a community-based case–control study of gout and attacks of self-limiting synovitis. From 4249 questionnaire respondents, 488 subjects who had indicated a history of gout or attacks of self-limiting synovitis were invited to attend a clinical assessment, 359 of whom agreed to attend.

Table I
Diagnoses of the 459 study participants

Diagnosis	*n*
Gout	167
Osteoarthritis	143
Rheumatoid arthritis	49
Other crystal arthritis	37
Seronegative spondarthropathy	17
Unspecified inflammatory arthritis	12
Tendinopathy	9
Mechanical back pain/fibromyalgia/chronic pain	9
Viral arthritis	5
Connective tissue disease	7
Asymptomatic hyperuricaemia	1
Peripheral oedema	1
Trauma	1
Extrinsic allergic alveolitis	1

Table II
Prevalence of grades of HV severity

Observer assessment	A/F	B/G	C/H	D/I	E/J	Total
Left foot	107 (23%)	216 (47%)	88 (19%)	28 (6%)	20 (4%)	459
Right foot	112 (24%)	203 (44%)	81 (18%)	36 (8%)	27 (6%)	459

Table III
Weighted kappa scores for five-grade HV scale

	Weighted kappa (95% CI)		
	Left foot	Right foot	Overall
Subject−observer agreement	0.44 (0.38, 0.50)	0.45 (0.40, 0.51)	0.45 (0.41, 0.49)
Subject repeatability (1−2 months)	0.46 (0.30, 0.61)	0.60 (0.45, 0.75)	0.53 (0.42, 0.64)
Subject repeatability (3−6 months)	0.49 (0.42, 0.57)	0.53 (0.45, 0.60)	0.51 (0.46, 0.56)
Observer repeatability	0.79 (0.63, 0.95)	0.84 (0.71, 0.98)	0.82 (0.71, 0.95)

These subjects had completed a postal questionnaire which included the self-report instrument. A face-to-face assessment including blinded observer assessment and repeated subject self-report was then undertaken in two groups: (1) in 288 respondents seen 3−6 months following the completion of the questionnaire; and (2) in 71 respondents seen after 1−2 months after completion of the questionnaire.

All assessments were undertaken by a single observer, a trainee rheumatologist of 4 years experience (ER).

ANALYSIS

Data from the two sources of recruitment for the study were combined and cross-tabulated. Validity of the instrument was assessed by the weighted kappa statistic and its 95% confidence interval (CI) for subject−observer agreement ($n = 459$). Reliability was assessed by the weighted kappa statistics (95% CI) for subject repeatability at 1−2 months ($n = 71$), subject repeatability at 3−6 months ($n = 313$) and observer repeatability ($n = 25$). Kappa statistics were interpreted as follows: <0.20, poor; 0.21−0.40, fair; 0.41−0.60, moderate; 0.61−0.80, good; and 0.81−1.00, very good[13]. Analysis was performed for left and right feet combined and then individually.

The HV grade was also dichotomised as present or absent by classifying the three most severe grades as present, specifically, ''A'' or ''B'' (left foot) or ''F'' or ''G'' (right foot) as absent and ''C'', ''D'' or ''E'' (left foot) or ''H'', ''I'' or ''J'' (right foot) as present. Non-weighted kappa and its 95% CI were calculated for subject−observer agreement, subject repeatability at 1−2 months and 3−6 months and observer repeatability as above. Taking observer assessment as the ''gold standard'', sensitivity and specificity of subject assessment using the dichotomised instrument were also calculated.

Since it was intended to use the self-report instrument in a subsequent study to assess the association between HV and great toe pain, it was important to ensure that HV severity was not over-reported in subjects with great toe symptoms. Subjects were classified as having great toe symptoms if they reported either a history of acute attacks of gout or chronic pain at the first MTP joint. These data were only available for community-derived subjects. Possible over-reporting of HV by subjects with great toe symptoms was assessed in two ways. Firstly, subject HV assessment on the five-grade scale was compared to observer assessment. Subject assessment was classified as ''over-reported'' when it was more severe than observer assessment, ''correct'' when it agreed with observer assessment and ''under-reported'' when it was less severe than observer assessment. Odds ratios (ORs) (95% CI) were calculated between great toe symptoms and reporting status (i.e., under-reported vs correct subject assessments and then over-reported vs correct subject assessments). Secondly, ORs (95% CI) were calculated between great toe symptoms and subject assessment and then observer assessment on the dichotomised scale. Indirect comparison of the ORs was made.

Analyses were performed using SPSS software version 11.0 apart from the weighted kappa statistic which was calculated at an internet-based statistics application (http://faculty.vassar.edu/lowry/VassarStats.html).

Results

Four hundred and fifty-nine subjects participated in the study. Two hundred and thirty-six (51%) were females. Participant's diagnoses are shown in Table I. Prevalence of the different grades of HV severity is shown in Table II.

Using the five-grade HV scale, subject−observer agreement and subject repeatability at both time points were moderate (Table III). Agreement was generally similar for left and right feet apart from subject repeatability at 1−2 months where agreement was better for the right foot than the left (weighted kappa 0.60 compared to 0.46), although there was considerable overlap of the 95% CIs. Observer repeatability was good for the left foot and very good for the right foot and overall, although numerically the weighted scores for left and right feet were only 0.05 apart.

Using the dichotomised scale, subject−observer agreement was moderate whereas subject repeatability at both time points was good (Table IV). Observer repeatability was very good. Again, agreement was similar for left and right feet apart from subject repeatability at 1−2 months where agreement for the right foot was better than the left (unweighted kappa 0.73 vs 0.54), although the 95% CIs overlapped. Sensitivity was 75% for the left foot, 76% for the right foot and 75% overall. Specificity was 82% for the left foot, 82% for the right foot and 82% overall.

Table IV
Kappa scores for dichotomised HV scale

	Kappa (95% CI)		
	Left foot	Right foot	Overall
Subject−observer agreement	0.54 (0.46, 0.62)	0.56 (0.48, 0.64)	0.55 (0.49, 0.61)
Subject repeatability (1−2 months)	0.54 (0.34, 0.75)	0.73 (0.56, 0.89)	0.63 (0.50, 0.77)
Subject repeatability (3−6 months)	0.61 (0.51, 0.71)	0.61 (0.51, 0.70)	0.61 (0.54, 0.68)
Observer repeatability	0.83 (0.60, 1.06)	0.83 (0.61, 1.05)	0.83 (0.67, 0.99)

Table V
OR between subject over-reporting of HV and great toe symptoms

	Subject HV reporting			OR (95% CI)	
	Under-reported	Correct	Over-reported	Under-report vs correct	Over-report vs correct
Great toe symptoms	50 (15%)	174 (53%)	107 (32%)	1.267 (0.840–1.911)	1.218 (0.870–1.706)
No great toe symptoms	75 (20%)	206 (54%)	104 (27%)		

Subject assessment of HV severity agreed exactly with observer assessment in just over half of the observations made in both subjects with and without great toe symptoms (Table V). Over-reporting was slightly more frequent in those with great toe symptoms and under-reporting was slightly more frequent in those without great toe symptoms but 95% CIs for the ORs crossed unity.

The presence of great toe symptoms was associated with subject assessment of HV in the left foot (OR 1.69; 95% CI 1.08, 2.64) and overall (OR 1.52; 95% CI 1.11, 2.07) but not in the right foot (Table VI). There was no association between great toe symptoms and observer assessment of HV. Although this suggests that subjects with great toe symptoms may over-report HV compared to observer, there was overlap of 95% CIs on indirect comparison of ORs for subject and observer assessment.

Discussion

This study describes the validation of the first instrument for subject self-reported HV in a large, mixed population of hospital and community-derived subjects. Self-reported HV assessed by this instrument, on either a five-grade or dichotomised scale, has been shown to be both valid and reliable and, as such, appears to be suitable for use in a large community-based epidemiological study. Kappa scores for subject–observer agreement and subject repeatability were numerically higher using the dichotomised scale compared to the five-grade scale. However, the difference was not statistically significant (the 95% CI overlapped) suggesting that the validities (subject–observer) and reliabilities (subject–subject) are similar. The numerical differences may be due to the greater agreement expected on a dichotomised scale given the smaller number of categories, which may not have been fully adjusted for by the weighted kappa statistics.

Most previous studies of HV have relied upon direct clinical assessment which is both costly and time-consuming. Alternatives include the Manchester scale, a set of four photographs, which has been validated for use in epidemiological studies but has not been validated for patient self-report[10]. Further advantages of the line drawings over the Manchester scale are a larger number of grades, depiction of both right and left feet, demonstrated subject repeatability at 3–6 months (compared to 2 weeks) and validation of an additional dichotomised definition of HV. The choice between the five-grade and dichotomised grading of HV allows flexibility according to study design and the precise research question being asked. The Manchester scale is also disadvantaged by the depiction of metatarsus primus varus in addition to HV in the most severe grade and the depiction of one left and three right feet in the original report, although this was corrected for the recent radiographic validation[12].

A number of caveats apply to this study. Some may question the biological validity of the line drawings which were developed from a single photograph of a normal foot with each grade representing a 15° increase in HV angle rather than a "real life" model of HV. In contrast, the Manchester scale consists of four photographs representing "no deformity", "mild deformity", "moderate deformity" and "severe deformity" chosen from a larger series of photographs and has been validated against the radiographic hallux abducto angle[12]. However, the sequential 15° increase incorporated into the line drawings allows a greater degree of precision of the severity of HV. A second caveat is that only a single trained observer was used to act as the gold standard for subject self-assessment, rather than radiographs or the Manchester scale, and has not been validated against another trained observer. Although the observer's assessments were shown to be highly repeatable, no measure of inter-observer agreement was sought. A further caveat is that the instrument has not been validated in normal subjects, which is an important consideration as it is intended for use in a community survey. However, although all of the subjects had musculoskeletal problems, in many of them the first MTP joint was not affected or HV was not present providing some evidence of wider generalisability. A fourth caveat concerns the analysis of data combining observations for left and right feet. Presented data for left and right feet are not truly independent and for an epidemiological study, individuals may be designated HV positive or negative according to whether either foot is affected by HV. However, for a validation study such as this one, it is appropriate to consider observations for left and right feet combined and separately. In view of the intention to use this self-report instrument to examine the association of HV and gout and great toe pain, a fifth caveat concerns the possibility that subjects with great toe symptoms might over-report the severity of HV. A small association was seen between subject assessment of HV and great toe pain whereas no association was seen between observer assessment and pain. However, the magnitude of the association between subject assessment was small and when individual subject observations were classified as under-reported, correct or over-reported against the observer assessment, no association between great toe symptoms and over-reporting of HV severity was found.

Table VI
OR between great toe symptoms and HV (≥grade C/H)

	OR (95% CI)		
	Left	Right	Overall
Subject assessment	1.692 (1.084, 2.639)	1.360 (0.878, 2.106)	1.518 (1.111, 2.073)
Observer assessment	1.319 (0.815, 2.133)	1.352 (0.807, 2.265)	1.318 (0.928, 1.871)

In summary, an instrument for patient self-reported HV consisting of a series of line drawings has been developed. Both the five-grade and dichotomised scales of HV appear to have validity and reliability for use in epidemiological studies.

Acknowledgements

We would like to thank the staff and patients of Arnold Health Centre, The Calverton Practice and the Rheumatology Clinic at Nottingham City Hospital, UK. We are grateful for funding from the Arthritis Research Campaign, UK (ICAC grant 14851; WZ Senior Lecturership) and unrestricted financial support from Astra-Zeneca-UK, Glaxo-Smith-Kline-USA and Ipsen.

References

1. Benvenuti F, Ferrucci L, Guralnik JM, Gangemi S, Baroni A. Foot pain and disability in older persons: an epidemiologic survey. J Am Geriatr Soc 1995;43:479–84.
2. Elton PJ, Sanderson SP. A chiropodial survey of elderly persons over 65 years in the community. Public Health 1986;100:219–22.
3. White EG, Mulley GP. Footcare for very elderly people: a community survey. Age Ageing 1989;18:276–8.
4. Dunn JE, Link CL, Felson DT, Crincoli MG, Keysor JJ, McKinlay JB. Prevalence of foot and ankle conditions in a multiethnic community sample of older adults. Am J Epidemiol 2004;159:491–8.
5. Leveille SG, Guralnik JM, Ferrucci L, Hirsch R, Simonsick E, Hochberg MC. Foot pain and disability in older women. Am J Epidemiol 1998;148:657–65.
6. Badlissi F, Dunn JE, Link CL, Keysor JJ, McKinlay JB, Felson DT. Foot musculoskeletal disorders, pain and foot-related functional limitation in older persons. J Am Geriatr Soc 2005;53:1029–33.
7. Garrow AP, Silman AJ, Macfarlane GJ. The Cheshire Foot Pain and Disability Survey: a population survey assessing prevalence and associations. Pain 2004; 110:378–84.
8. Menz HB, Lord SR. The contribution of foot problems to mobility impairment and falls in community-dwelling older people. J Am Geriatr Soc 2001;49:1651–6.
9. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. N Engl J Med 1988;319:1701–7.
10. Garrow AP, Papageorgiou A, Silman AJ, Thomas E, Jayson MI, Macfarlane GJ. The grading of hallux valgus. The Manchester Scale. J Am Podiatr Med Assoc 2001;91:74–8.
11. Menz HB, Tiedemann A, Kwan MM, Latt MD, Sherrington C, Lord SR. Reliability of clinical tests of foot and ankle characteristics in older people. J Am Podiatr Med Assoc 2003;93:380–7.
12. Menz HB, Munteanu SE. Radiographic validation of the Manchester scale for the classification of hallux valgus deformity. Rheumatology (Oxford) 2005;44:1061–6.
13. Landis JR, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33: 159–74.