

Concise report

The reliability, validity and responsiveness of an aggregated locomotor function (ALF) score in patients with osteoarthritis of the knee

C. J. McCarthy and J. A. Oldham

Objectives. The aggregated locomotor function (ALF) score, a simple measure of observed locomotor function, using timed walking, stairs and transfers, was developed and evaluated for intra-tester reliability, criterion-related validity and responsiveness in a sample of patients with knee osteoarthritis.

Methods. Patients with knee osteoarthritis ($n = 214$) were recruited for inclusion in a randomized controlled trial investigating two methods of exercise provision. Before treatment, patients completed the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and Short Form 36 health survey (SF-36) questionnaires and were timed whilst performing an 8 m walk, ascending and descending a set of gymnasium stairs and completing a test of transferring in and out of a chair. A group of 15 patients also undertook a replicate test–retest reliability study of the above outcome measures. Standardized response means were calculated for the ALF, WOMAC and SF-36 from data from the clinical trial.

Results. The ALF takes 10 min to administer and demonstrated excellent intra-tester reliability, with excellent intra-class correlation coefficient (ICC) statistics (ICC_{2,k} 0.99; 95% CI 0.98–0.99), and low standard error of measurement (0.86 s) and smallest detectable difference (9.5%) values. Criterion-related validity with the physical function dimensions of the WOMAC and SF-36 was good, with correlation coefficients of 0.59 and –0.53 respectively. Standardized response means were higher for the ALF (0.49) than for both the WOMAC (0.39) and the SF-36 (0.12).

Conclusions. This work has demonstrated that the ALF can be used as a measure of physical function status and as a means of quantifying treatment response. The measure offers a simple and convenient outcome in the assessment and treatment of locomotor dysfunction. The ALF score is a reliable, valid and responsive outcome measure over 12 months and can be recommended for use in the evaluation of patients with knee osteoarthritis.

KEY WORDS: Locomotor function, Osteoarthritis, Validity, Reliability.

One of the main aims of physiotherapeutic treatment of chronic disabling conditions, such as knee osteoarthritis (OA), is to improve locomotor function. To evaluate improvement of locomotor function, clinically convenient and valid outcome measures are required. The disability experienced by patients with OA of the knee has been assessed in many ways. Two common approaches are the use of self-evaluation questionnaires such as the Western Ontario and McMaster Universities OA index (WOMAC) [1] and the use of performance observation [2], such as timed walking tests [3] and the ‘timed up and go’ test [4]. Both methods have their own specific advantages and disadvantages. Observational methods have been shown to demonstrate good criterion-related validity with self-evaluation methods of disability measurement, particularly when assessing mobility [5], and are considered to provide measures that are less influenced by patient expectation of treatment effect [6].

Objective assessment of locomotor function of timed walking, stair ascent and descent, and transferring to and from a chair has been used by several investigators in the field of knee osteoarthritis [7–10]. Recently the times of these individual activities have been aggregated to form one timed score, with the rationale that ‘any single test imparts little information about the patient’s overall

functional ability, and that by aggregating the time of the activities a better objective assessment of the patient’s overall functional capabilities can be obtained’ [9].

Whilst the individual locomotor functions of walking [11], stair ascent and descent [12] and transferring from sitting to standing [4] have established validity, the reliability, validity and responsiveness of an aggregated locomotor function score, comprising these three functions, has not been established.

Methods

This study comprised two parts. A test–retest reliability study performed with a small group of patients ($n = 15$) was followed by an evaluation of the criterion-related validity of the ALF score with a large study sample ($n = 214$).

Patients

Two hundred and fourteen patients with knee OA were recruited for inclusion in a randomized controlled trial (RCT) investigating

Centre for Rehabilitation Science, Manchester Royal Infirmary, Oxford Road, Manchester M13 9WL, UK

Submitted 7 July 2003; revised version accepted 22 October 2003.

Correspondence to: C. J. McCarthy. E-mail: christopher.j.mccarthy@man.ac.uk

the long-term effect of two methods of exercise provision [13]. Patients were included if they met the American College of Rheumatology (ACR) clinical diagnostic guidelines for knee osteoarthritis [14, 15] and had radiological evidence of OA. Patients were excluded if they had knee OA secondary to inflammatory arthritis, had significant psychiatric or general medical morbidity that would either preclude their undertaking the exercises or their understanding of the nature of the exercise treatment, or had received an intra-articular steroid injection in the knee within 3 months. A convenience sample of 15 knee OA patients, meeting the same selection criteria, was used for a test-retest reliability study. This study had received approval from the Central Manchester Local Research Ethics Committee and informed, written consent was obtained from all patients according to the Declaration of Helsinki.

Reliability study

A reliability study was conducted with a group of knee OA patients ($n=15$). Patients attended for assessment and then undertook a replicate assessment within 1 week. The group undertook measurements of walking, stair ascent/descent and transferring, and completed the WOMAC and Short Form 36 health survey (SF-36) questionnaires at both assessments.

Outcome measures

The aggregated locomotor function (ALF) score. This outcome score was formed by summing the mean timed scores (seconds) from three locomotor functions (walking time, stair ascent and descent time) and time taken to transfer from sitting to standing.

Eight metre walk time. Patients were asked to walk, at their own naturally preferred 'comfortable' pace, across the floor of the physiotherapy gymnasium. Following recommendations [16], a 10 m stretch of floor was used. An 8 m distance was marked on the gymnasium floor. Timing of the central 8 m allowed one or two steps at either end of the walk for untimed acceleration and deceleration, a process that has been shown to increase test-retest reliability [16]. The time (s) taken to complete the distance was measured using a hand-held stopwatch (Zeon, UK). Patients were permitted to use walking aids if they required them. Three repetitions of the walk were undertaken and the times recorded. The mean of times was calculated and used for subsequent analysis.

Stair ascent and descent time. Patients were asked to ascend and then descend seven steps (four of 15 cm and three of 20 cm). Patients were instructed to undertake this task at their naturally preferred comfortable pace. The method that the patient employed to negotiate the stairs was recorded, i.e. whether they used alternate legs, used the banisters or always led with one leg. Patients were permitted to use the two banisters if they felt it necessary, as the use of banisters has been shown to not affect times [17]. Patients were timed (in seconds) using a hand-held stopwatch and repeated the test four times. The mean of the four repetitions was calculated and used for subsequent analysis. Four repetitions were used as the stairs used had steps of different heights; thus, by going over the steps in one direction and then the other the patients ascended and descended the different height steps twice.

Transferring time. Patients were asked to walk, at their own natural pace, a distance of 2 m to a chair and sit down, then immediately stand up and walk back to the start. Patients were timed (in seconds) using a hand-held stopwatch as they approached and retreated from the chair. The chair had no arms and a seat height of 0.46 m, typical of a toilet seat height [18].

Patients undertook three timed repetitions, the mean of which was calculated and used for subsequent analysis.

WOMAC. This is a tri-dimensional, disease-specific, self-administered health status measure [19]. The Likert scale version (LK3.0) of the questionnaire was used. Missing data and scoring procedures followed the WOMAC user guidelines [1].

SF-36. This is a 36-item questionnaire which measures health functioning on eight scales, including a physical functioning scale, and is among the most widely used measures of quality of life in studies of patients and populations [20]. Missing data and scoring procedures followed the SF-36 user guidelines [21].

Analysis

Test-retest reliability. Data from these two assessments was then analysed for intra-rater reliability. To calculate useful indices of reliability, four statistics were calculated: intra-class correlation coefficients ($ICC_{2,k}$) with 95% confidence intervals (CI), standard errors of measurement (SEM) and smallest detectable differences (SDD). The SDD is derived from the SEM and can be expressed in original units or as a percentage of the feature's grand mean. This statistic indicates the level of change (or percentage change) in a feature attributed with 95% certainty to a true change in the condition of the subject instead of that being caused by test-retest errors [22].

Criterion-related validity analysis. To evaluate the strengths of the correlations between the 'new' outcome measure and the two validated measures, the individual components and total ALF scores were correlated with the physical function dimensions of the WOMAC and SF-36 indexes, obtained at the subject's first assessment ($n=214$). Correlations coefficients of around 0.2 were considered small, over 0.5 as moderate and over 0.8 as large [23].

Responsiveness analysis. The between-group standardized response means (SRM), obtained from the long-term review of a RCT evaluating the relative effectiveness of two methods of providing exercise, were calculated and compared. The SRM was calculated by dividing the mean change score of the measure by the standard deviation of the change scores over the 12 month time period [24]. A larger SRM for an outcome measure would suggest greater responsiveness to the intervention by that outcome.

Results

The demographic and baseline data for patients are presented in Table 1. Table 2 shows the reliability for each of the locomotor function times and also for the ALF score. The ALF demonstrated an extremely low SDD score (9.5%) and a high ICC statistic (0.99) with a narrow confidence interval (0.98 to 0.99).

Due to non-normal distribution, Spearman's rank correlations between the ALF score and the physical function dimension of the WOMAC and SF-36 were calculated. Moderately sized, statistically significant correlations were demonstrated with the WOMAC ($r_s=0.59$) and SF-36 ($r_s=-0.53$).

The standardized response means obtained from the RCT of exercise provision at 12 month follow-up were larger for the ALF (SRM = 0.49) than the physical function domains of the WOMAC (SRM = 0.39) and SF-36 (SRM = 0.12) (Friedman test $\chi^2=15.5$, $P<0.001$).

Discussion

The ALF score has demonstrated excellent intra-tester reliability with a high ICC statistic, narrow confidence intervals, low SEM and low SDD. Importantly, the measure demonstrates moderately sized correlation with two validated self-report questionnaires of physical function and appears to be more responsive to change, induced by exercise intervention, than either.

TABLE 1. The demographic and baseline patient data from the reliability and validity study samples

	Reliability study (<i>n</i> = 15)		Validity study (<i>n</i> = 214)	
	Mean	S.D.	Mean	S.D.
Age (yr)	65.1	11.3	64.7	9.8
BMI (kg/m ²)	31.1	4.4	29.8	5.3
Symptom duration (yr)	9.3	8.4	7.9	7.3
WOMAC				
Pain domain	9.0	3.8	10.0	3.7
Physical function	29.1	14.4	30.2	14.0
SF-36 physical function	37.3	22.6	36.5	23.2
Walk time (s)	7.7	1.8	8.6	4.8
Stair time (s)	8.6	4.9	8.8	5.3
Transfers (s)	7.4	2.2	8.3	3.9
ALF (s)	23.7	8.4	25.5	14.0

BMI, body mass index.

TABLE 2. Reliability indices and correlation coefficients

	ICC _(2,k) (<i>n</i> = 15)	95% CI	SEM (<i>n</i> = 15)	SDD (%) (<i>n</i> = 15)	WOMAC (<i>r_s</i>) (<i>n</i> = 214)	SF-36 (<i>r_s</i>) (<i>n</i> = 214)
Walk time	0.98	0.97 to 0.99	0.26 (s)	18.4	0.57	-0.44
Stair time	0.98	0.95 to 0.99	0.87 (s)	30.6	0.58	-0.41
Transfers	0.99	0.96 to 0.99	0.38 (s)	22.7	0.57	-0.41
ALF	0.99	0.98 to 0.99	0.86 (s)	9.5	0.59	-0.53

The Spearman's rank correlation coefficients (*r_s*) are between the physical function dimensions of the WOMAC and SF-36 and the ALF score. All values *P* < 0.01.

If valid outcome measures are to be used in clinical practice, they must be appropriate for the patient population and convenient for the clinician. This work has demonstrated that a simple timed measure of locomotor function can be used as a measure of physical functioning and as a means of quantifying treatment response. The individual components of the ALF challenge the locomotor function of patients with knee OA, but are not so demanding that they cannot be completed. Consequently, the measure appears to offer the patient and clinician an appropriate, simple, quick (time to administer 10 min) and convenient outcome measure in the treatment of knee OA.

The use of the ALF score can be recommended for consideration as a clinical and research outcome measure with knee OA patients. Based on the evidence presented above, the ALF offers a valid, simple and convenient outcome measure which is responsive over 12 months and can be considered when planning treatment evaluation with OA patients.

The authors have declared no conflicts of interest.

Acknowledgements

This project was funded by the NHS Health Technology Assessment Agency (reference 94/39/14). The views and opinions expressed in this report do not necessarily reflect those of the NHS Executive.

Rheumatology	Key message
	The aggregated locomotor function (ALF) score is a convenient and simple timed outcome measure. It is reliable, valid and responsive over 12 months to change in patients with knee osteoarthritis.

References

- Bellamy N. Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index: Users Guide. London, Ontario: University of Western Ontario, 1996.
- Wijlhuizen GJ, Ooijendijk W. Measuring disability, the agreement between self evaluation and observation of performance. *Disabil Rehabil* 1999;21:61-7.
- Haghani H, Marks R. Relationship between maximal isometric knee extensor and flexor strength measures, age and walking speed of healthy men and women ages 18-74. *Physiother Can* 2000;52:33-8.
- Podsiadlo D, Richardson S. The timed 'Up and Go': a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991;39:142-8.
- Steultjens MP, Dekker J, van Baar ME, Oostendorp RA, Bijlsma JW. Internal consistency and validity of an observational method for assessing disability in mobility in patients with osteoarthritis. *Arthritis Care Res* 1999;12:19-25.
- Watson D, Pennebaker JW. Health complaints, stress and distress: exploring the central role of negative affectivity. *Psychol Rev* 1989;96:234-54.
- Ettlinger WH Jr, Burns R, Messier SP *et al*. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). *JAMA* 1997;277:25-31.
- Hopman-Rock M, Westhoff MH. The effects of a health educational and exercise program for older adults with osteoarthritis of the hip or knee. *J Rheumatol* 2000;27:1947-54.
- Hurley MV, Scott DL. Improvements in quadriceps sensorimotor function and disability of patients with knee osteoarthritis following a clinically practicable exercise regime. *Br J Rheumatol* 1998;37:1181-7.
- Van Baar ME, Dekker J, Oostendorp RA *et al*. The effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a randomized clinical trial. *J Rheumatol* 1998; 25:2432-9.
- Kerrigan DC, Todd MK, Della Croce U, Lipsitz LA, Collins JJ. Biomechanical gait alterations independent of speed in the healthy elderly: evidence for specific limiting impairments. *Arch Phys Med Rehabil* 1998;79:317-22.

12. Jette AM, Jette DU, Ng J, Plotkin DJ, Bach MA. Are performance-based measures sufficiently reliable for use in multicenter trials? Musculoskeletal Impairment (MSI) Study Group. *J Gerontol A Biol Sci Med Sci* 1999;54:M3-6.
13. McCarthy CJ, Pullen R, Mills PM, Roberts C, Silman AJ, Oldham JA. Supplementing home exercise with class-based exercise leads to reductions in pain in knee osteoarthritis, but no greater muscle strength or compliance with home exercise at long term follow-up. *Rheumatol* 2003;42 (Suppl. 1):17-18.
14. Altman RD, Asch E, Bloch DA *et al*. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis Rheum* 1986; 29:1039-49.
15. Altman RD. Criteria for classification of clinical osteoarthritis. *J Rheumatol* 1991;18(Suppl 27):10-12.
16. Hirokawa S. Normal gait characteristics under temporal and distance constraints. *J Biomech Eng* 1989;11:449-56.
17. Bassey EJ, Fiatarone MA, O'Neill EF, Kelly M, Evans WJ, Lipsitz LA. Leg extension power and functional performance in very old men and women. *Clin Sci* 1992;82:321-7.
18. Howe TE, Oldham JA. Functional tests in elderly osteoarthritic patients: variability of performance. *Nurs Standard* 1995;9:35-8.
19. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833-40.
20. Hemingway H, Stafford M, Stansfield S, Shipley M, Marmot M. Is the SF-36 a valid measure of change in population health? Results from the Whitehall II study. *Br Med J* 1997;315:1273-9.
21. Ware JE Jr, Snow K, Kosinski M, Gandek B. SF-36 health survey manual and interpretation guide. Boston, MA: New England Medical Centre, 1993.
22. Sim J, Wright C. Research in healthcare. concepts, designs and methods. Cheltenham, UK: Stanley Thornes, 2000.
23. Cohen J. A power primer. *Psychol Bull* 1992;112:155-9.
24. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459-68.