The test/retest reliability of physical performance measures in osteoarthritis patients

Objectives: There is limited data examining the reliability and consistency of objective measures of osteoarthritis (OA) function. Therefore the aim of this study was to assess their test/retest reliability.

Methods: 21 postmenopausal women over the age of 50 diagnosed with bilateral knee OA were recruited from a single centre. Endpoints assessed included time to walk 20 m, 40 m, ascend and descend 11 stairs, pain scores, standard WOMAC questionnaires and range of motion of the knee with assessments repeated at a follow up visit.

Results: Time to walk 20 m, 40 m, and climb and descend stairs and total WOMAC scores between the two visits was highly reliable. There was no reliability for the range of motion of the knee.

Conclusions: The high degree of reliability for the endpoints of walking and stair climbing within patients and across visits supports the use of these as outcome measures within clinical trials of OA.

Keywords: osteoarthritis • knee • performance measures

Frances C Humby[*1]**, Alison J Hughes**[2] **& Ali S M Jawad**[1]

[1]Department of Rheumatology, Mile End Hospital, Barts Health NHS Trust, London, United Kingdom

[2]Glaxo Smith Kline Consumer Healthcare St. Georges Avenue, Weybridge, Surrey, United Kingdom

*Author for correspondence:
Ali.jawad@bartshealth.nhs.uk

Introduction

Osteoarthritis (OA) is the most common musculoskeletal disorder and is strongly age related, with worldwide prevalence rates estimated to be approximately 10% for men and 18% for women between the ages of 65-74 [1]. It is characterised clinically by pain localized around the affected joint(s), limitation of movement and loss of function and is often accompanied by changes in physical function such as the ability to walk on a flat surface and to climb stairs. In all joints except the hip, it is more common in women than in men [1]. OA is slow to evolve, but often has periods of relative stability lasting many years but is equally marked by periods of flare followed by remission [2]. Although the flares may be precipitated by specific factors such as trauma, often there is no apparent cause. The precise cause of OA is unknown, but it is primarily a degenerative disorder affecting the articular cartilage of weight bearing joints. There is an associated disability with worsening OA, causing lost days from work and early retirement.

Changes in physical function that are documented with OA do not solely relate to pathological changes accompanying the disease such as cartilage degeneration, sub-chondral bone stiffening and active new bone formation. Instead avoidance of and impairment of physical function may also be related to pain and the fear of pain. Consistent with the importance of determining physical function the gold standard questionnaire methods of assessing the severity of OA disease such as the WOMAC(Western Ontario and McMasters University Osteoarthritis Index) [3] consist of a significant recall based measure of everyday physical function. These questionnaire based measures have demonstrated sensitivity in detecting the efficacy of analgesics such as non-steroidal anti-inflammatory drugs [3]. Whilst recall based questions are extremely valuable there is also increasing data to suggest that objective measures are of equal importance in clinical studies aimed at determining the efficacy of OA treatments [4-7]. However although there are published data on subjective assessment scales for function of OA of the knee [8] which assess relevance and reproducibility

and data examining objective functional assessments for example walk time, stair climb time or time to stand from sitting the reliability of these objective measures in OA patients has not been extensively investigated [9], a critically important step to address before such outcomes are routinely integrated as outcome measures into clinical trial protocols investigating OA. Therefore the aim of this study was to explore the reliability of objective functional measures of OA function.

Methods

Patients

21 patients were recruited from the general rheumatology clinic, Department of Rheumatology Mile End Hospital, Barts Health NHS Trust. 20 patients completed the study. All patients were postmenopausal women over the age of 50 years and were diagnosed with bilateral knee OA. Patients with other comorbidities that may have affected pain assessments such as rheumatoid arthritis or fibromyalgia as determined by the clinician were excluded. All patients were recruited following written informed consent and following the approval of the local ethics committee (East London and City Research Ethics Committee).

Study design

All patients attended for a screening visit followed by 2 assessment visits (range 2-14 days). All OA treatments (including non-pharmacological treatments) were stopped for 2 weeks prior to the screening visit, patients were allowed to dose with paracetamol for up to 24 hours prior to the assessment visits. No intra articular steroids were permitted for 3 weeks prior to study inclusion. Assessments were made in the morning at the same time of day for each patient for each visit. No exercise was permitted for 12 hours prior to each assessment visit and transport to the hospital was provided to ensure this was adhered to. Walking aids normally used were permitted during assessment visits.

At each patient visit the following assessments were performed:

- Subjective pain assessment on a 5 point Likert scale.
- Abbreviated WOMAC scale: The patient's difficulty in performing various activities as detailed in the abbreviated WOMAC scale [8] (ascending stairs, sitting, walking on the flat, getting into/

out of a car, putting on socks/stockings, rising from bed and rising from sitting) were rated by patients on a 5 point Likert scale.

- Walk time: Patients were timed by the investigator to walk 40 m per assessment visit, the time was also recorded for the patient to walk the first 20 m.
- Stair climbing: The time for patients to climb and descend 11 steps with a handrail was recorded.
- Time to stand from sitting: Patients seated on a hard backed chair without arms were timed from initiation of movement to time to standing and returning to sitting position and,
- Range of knee motion: Patients had the range of knee motion (maximum and minimum angles of both left and right knee) measured by a goniometer (ADU301 Biometrics Gwent, UK) by a single investigator.

Statistical Analysis

The between-patient and within-patient variability were assessed from the analysis of variance on each measurement with patient as an effect in the model and reported as $R2$ presented as the percentage of each adding to 100. The variability and consistency of the measures was also reported as mean, median, range and standard deviation. A weighted Kappa was calculated to assess agreement (reliability) in measure between the two visits.

Results

Patients

All 21 patients recruited to the study were female. 17 (81%) were Caucasian, 2 (9.5%) black and 2 (9.5%) Asian. The mean age was 66.8 years (SD 9.6, range 50.8-87). One patient withdrew consent before study completion and was not included within the analysis.

Assessments

Subjective assessment of pain shows low re-test reliability

In order to determine whether a subjective measure of pain could be used as a reliable endpoint following intervention in clinical trials, Likert scores for pain performed at the baseline and follow up assessment visits were compared. There was a low correlation Table 1 between

Table 1. Summary of variability between individual outcome measures between first and second assessments.					R2 (%)		Kappa (between visits)
		Visit 1 Mean (SD)	Visit 2 Mean (SD)	Difference (visit 2-visit 1)	Between subject	Within subject	
Total WOMAC Score		10.39 (7.9)	10.44 (8.09)	0.06 (3.73)	94.6	5.4	0.69
Stair climbing time (s)		31.8 (28.7)	33.95 (32.4)	2.15 (11.4)	96.4	3.62	0.79
Stand from sitting time (s)		4.85 (3.2)	5.4 (3.3)	0.55 (2.46)	85.1	14.9	0.54
Time to walk 20 m (s)		26.8 (15.4)	27.8 (20.8)	1 (8.12)	95	5.01	0.78
Time to walk 40 m (s)		53.5 (31.2)	55.3 (40.2)	1.8 (15.97)	95	4.98	0.83
ROM knee	R	53.7 (20.2)	42.5 (24.4)	-11.2 (31.0)	48.9	51.1	0.03
	L	48.1 (18.7)	39.9 (23.5)	-8.2 (27.4)	56.3	43.7	0.11
Subjective assessment of pain	Beginning of day	1.2 (1.3)	1.6 (1.3)	0.4 (1.3)	74.1	26	0.38
	End of day	1.6 (1.43)	1.75 (1.41)	0.15 (1.04)	86.3	13.7	0.58
ROM- range of motion, R- right, L- left, s- seconds							

scores (weighted Kappa beginning of day 0.38 and end of day 0.58) suggesting that this would not be a reliable method to determine effect of intervention in studies of OA.

Time to walk 20 m/40 m showed a high level of re-test reliability

There was a marked difference in the time it took subjects to perform the 20 m walk ranging from 12 s for the fastest subject visit and 91 s for the slowest subject visit. The median time to perform the walk was 23 s on visit 1 and 21.5 s on visit 2. The median time taken to walk 40 m was 45 s during visit 1 and 42.5 s on visit 2. As with the 20 m walk there was a wide range of times taken to perform the walk ranging from 23 s on the fastest subject visit and 179 s on the slowest subject visit. For both the 20 m and 40 m walk there was a within subject variability of 5% and between subject variability of 95% Table 1. Weighted Kappa between visits was 0.78 and 0.83 respectively demonstrating good correlation between visits Table 1 and therefore a high level of re test reliability.

Time to climb and descend a set of 11 steps was highly reliable and reproducible between the two separate visits

The time to climb and descend the 11 stairs varied from 11 s for the fastest subject visit and 130 s for the slowest subject visit. The median for both visits was 22.5 s. As with the 20 m and 40 m walks, time to ascend and descend 11 steps was highly reliable and reproducible between the two visits with a within subject variability of 3.6%, a between subject variability of 96.4% Table 1 and a weighted Kappa of 0.79.

Time to stand up from sitting was more variable than other functional measures but still relatively

consistent between the two visits

The time to stand from sitting varied from 1 s for the fastest subject visit to 15 s for the slowest subject visit. The median for the first and second visits was 4.5 s and 5 s respectively. Compared to the other functional measures the time to stand up from sitting within subjects was more variable. The within subject variability was reasonably low at 14.9% compared to the between subject variability of 85.1% the weighted kappa was also lower at 0.54 Table 1.

Knee angle measurements were highly variable

Of all the parameters measured, the knee angle measurements obtained using the digital goniometers were the most variable. The within subject variability for the maximum angle of knee motion for the left knee was 43.7% compared to a between subject variability of 56.3% *versus* 51.1% and 48.9% for the right knee. The kappa was 0.11 and 0.03 respectively. Combined with the fact that the minimum angle was always zero, this data suggest that knee motion range as applied within this cohort would not be a reliable outcome measure in clinical trials.

WOMAC scores

The total WOMAC scores showed reliability similar to time to walk 20 m/40 m or climb stairs. However individual components of the WOMAC score such as ascending stairs, sitting and rising from sitting showed lower reliability Table 1. Overall however this confirms the usefulness of the well validated WOMAC score as an outcome measure for clinical trials of OA interventions.

Discussion

The significant health economic burden of OA

has led to recent advances in novel surgical and non-surgical approaches to treatment. However, the evaluation of such treatments requires robust outcome measures to determine efficacy. Although there is extensive data validating subjective measures of OA of the knee there is currently only limited data examining objective measures which are likely to be of equal importance when assessing clinical intervention for OA [10,11]. Previously published data support the reliability of the 40 m walk in assessing patients with hip OA [12] and the data presented herein support its application to patients with knee OA. Additionally a 20 m walk as well as an 11 step stair climb appears highly reliable and reproducible assessments when examining knee OA. The time to stand from sitting was the least reliable of the objective tests and this may be at least partly due to the difficulty in accurately timing the fastest of the subjects (i.e. those requiring less than 2 s to perform the task). Previous studies have also suggested that this may not be a reliable measure of OA function [13] and have not recommended it for use in patients with either hip or knee OA [9]. The current data would also support this recommendation.

The lack of reliability for the goniometer, used to measure the range of knee movement is likely due to methodological issues within this study. The most likely reason for this variability is experimental error as it was difficult to ensure that the transducers were affixed to the exact same site between visits. Furthermore it was also difficult to ensure that the elderly subjects consistently flexed and extended their legs to the same degree during the different visits. Importantly there is data to suggest use of a goniometer maybe a reliable measure of knee restriction in a previously published though smaller patient cohort [14]. Therefore with better training of both the operator as well as the subjects, this could still be a valuable objective measure of OA function and warrants further investigation.

Limitations of this current cohort are the inclusion of only women and a relatively small sample size. However the reliability of the results is supported by the demonstrated consistency in repeated measures of the extensively validated WOMAC score. However, the WOMAC has been criticized as there is little evidence to support the measurement properties of the stiffness subscale, and its test-retest reliability has been

low. In addition, studies have found inadequate factorial validity of the pain and physical function subscales, potentially weakening the ability of the physical function subscale to detect change when there is a weak association between pain and function [15].

Benefits of our data include the robust inclusion criteria and the stringent measures to ensure consistency between subject visits making translation of results to other patient cohorts realistic.

In conclusion, the data presented supports the conclusion that objective measurements of function, in particular the 20 m, 40 m walk and 11 step stair climb may be useful for evaluating the efficacy of OA treatments and may complement gold standard questionnaire assessments such as the WOMAC. However it is yet unclear how differences in the time taken to perform these physical functions relate to specific symptoms such as the degree of knee pain. At present it is also unclear how sensitive changes in these subjective end points might be to treatments of knee OA including analgesic use.

Conflict of interest

The authors declare they have no conflicts of interest.

References

1. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. *Bull. World. Health. Organ.* 81, 646–656 (2003).

2. Pavelka K. Symptomatic treatment of osteoarthritis: paracetamol or NSAIDs? *Int. J. Clin. Pract. Suppl.* 144, 5–12 (2004).

3. Bellamy N, Buchanan WW, Goldsmith CH *et al.* 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.* 15, 1833–1840 (1998).

4. Davis AM. Osteoarthritis year in review: rehabilitation and outcomes. *Osteoarthritis. Cartilage.* 20, 201–206 (2012).

5. Stratford PW, Kennedy DM, Riddle DL. New study design evaluated the validity of measures to assess change after hip or knee arthroplasty. *J. Clin. Epidemiol.* 62, 347–352 (2009).

6. Stratford PW, Kennedy DM, Woodhouse LJ. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee. *Phys. Ther.* 86,1489–1496 (2006).

7. Mizner RL, Petterson SC, Clements KE *et al.* Measuring functional improvement after total knee arthroplasty requires both performance-based and patient-report assessments: a longitudinal analysis of outcomes. *J. Arthroplasty.* 26, 728–737 (2011).

8. Whitehouse SL, Lingard EA, Katz JN *et al*. Development and testing of a reduced WOMAC function scale. *J. Bone. Joint. Surg. Br.* 85, 706–711 (2003).

9. Dobson F, Hinman RS, Hall M *et al*. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. *Osteoarthritis. Cartilage*. 20, 1548–1562 (2012).

10. Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J. Clinical. Epidemiol.* 59, 160–167 (2006).

11. Jordan KP, Wilkie R, Muller S *et al*. Measurement of change in function and disability in osteoarthritis: current approaches and future challenges. *Curr. Opin. Rheumatol.* 21, 525–530 (2009).

12. Wright AA, Cook CE, Baxter GD *et al*. A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis. *J. Orthop. Sports. Phys. Ther*. 41, 319–327 (2011).

13. Piva SR, Fitzgerald GK, Irrgang JJ. Get up and go test in patients with knee osteoarthritis. *Arch. Phys. Med. Rehabil*. 85, 284–289 (2004).

14. Brosseau L, Balmer S, Tousignant M *et al.* Intra and inter tester reliability and criterion validity of the parallelogram and universal goniometers for measuring maximum active knee flexion and extension of patients with knee restrictions. *Arch. Phys. Med. Rehabil*. 82, 396–402 (2001).

15. American College of Rheumatology. "Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)-General Description". *ACR* (2018).