

EXTENDED REPORT

Using patients' and rheumatologists' opinions to specify a short form of the WOMAC function subscale

F Tubach, G Baron, B Falissard, I Logeart, M Dougados, N Bellamy, P Ravaud

Ann Rheum Dis 2005;64:75–79. doi: 10.1136/ard.2003.019539

See end of article for authors' affiliations

Correspondence to:
Dr Florence Tubach,
Département
d'Epidémiologie,
Biostatistique et Recherche
Clinique, INSERM E0357,
Hôpital Bichat, 46 rue
Henri Huchard, 75018
Paris, France;
florence.tubach@
bch.ap-hop-paris.fr

Accepted 26 March 2004

Background: The WOMAC (Western Ontario and McMaster Universities) function subscale is widely used in clinical trials of hip and knee osteoarthritis. Reducing the number of items of the subscale would enhance efficiency and compliance, particularly for use in clinical practice applications.

Objective: To develop a short form of the WOMAC function subscale based on patients' and experts' opinions (WOMAC function short form).

Methods: WOMAC function subscale data (Likert version) were obtained from 1218 outpatients with painful hip or knee osteoarthritis. These patients and their rheumatologists selected the five items that they considered most in need of improvement. The rheumatologists were asked to select the five items for which patients in general are the most impaired. Items that were least important to patients and experts, those with a high proportion of missing data, and those with a response distribution showing a floor or ceiling response were excluded, along with one of a pair of items with a correlation coefficient >0.75 .

Results: The WOMAC function short form included items 1, 2, 3, 6, 7, 8, 9, and 15 of the long form. The short form did not differ substantially from the long form in responsiveness (standardised response mean of 0.84 v 0.80).

Conclusions: A short form of the WOMAC function subscale was developed according to the views of patients and rheumatologists, based on the responses of 1218 patients and 399 rheumatologists. The clinical relevance and applicability of this WOMAC function subscale short form require further evaluation.

One of the major uses of health measurement scales is detecting health status change over time, either in the context of clinical trials or epidemiological studies or as a strategy for monitoring the outcomes and making decisions about the care of individual patients in daily clinical practice. In all situations, a priority may be efficiency, achieved by the shortest possible questionnaire.¹ To date, methods of shortening questionnaires have focused on approaches that maintain the greatest internal consistency.² However, in the context of health measurement scales targeting a relatively heterogeneous disorder, it may be advantageous to sacrifice internal consistency for content validity.³

The Western Ontario and McMaster Universities (WOMAC) osteoarthritis index is a valid, reliable, and responsive measure in hip and knee osteoarthritis.^{4,5} This index is self administered and involves 17 items addressing the degree of difficulty in accomplishing 17 activities of daily life. While the mean importance score of the 17 items is similar at a group level, there is interindividual variability in the importance attached by individual patients to particular items.^{4,5} The WOMAC function subscale is short, and can be completed quickly. Nevertheless, an even shorter version would further enhance its applicability in epidemiological studies and for use in routine clinical practice.²

Our aim in this study was to specify a short form of the WOMAC function subscale dedicated to all patients with hip or knee osteoarthritis, by preserving the most important items for patients and rheumatologists (WOMAC function subscale short form).

METHODS

Study population

We conducted a prospective cohort study of four weeks' duration, involving 1362 outpatients with hip or knee osteoarthritis as defined by the American College of

Rheumatology,^{6,7} and 399 private rheumatologists in France. Each rheumatologist was required to include four patients, three with knee osteoarthritis and one with hip osteoarthritis. To be included in the study, patients had to experience pain from the osteoarthritis (≥ 30 mm on a visual analogue scale (VAS) ranging from 0 to 100 mm) and to require treatment with a non-steroidal anti-inflammatory drug (NSAID). Inclusion could begin with the onset of treatment or with a switch from one NSAID to another. Patients were excluded if they had a prosthesis on the assessed joint or if they had been treated with intra-articular injection in the four weeks before the study began. All patients initially visited the rheumatologist in charge of their case, and an NSAID was prescribed (the drug and its dosage were chosen by the physician). A final visit to the same rheumatologist was scheduled four weeks later.

Measurements

Patients and rheumatologists assessed the patient's status with respect to the osteoarthritis at the baseline visit and at week 4. Patients completed the French Canadian version of the WOMAC physical function subscale⁸ (17 items, five point Likert scale version, total score varying between 0 and 68; high scores indicate a high degree of functional impairment).

Patients were also asked to select the five items of the WOMAC function subscale that they considered most in need of improvement.

The rheumatologists were asked on one occasion to select the five items on the subscale which they consider result in the greatest impairment in patients with knee and hip osteoarthritis (not the specific patients they had included in the study).

Abbreviations: ICC, intraclass correlation coefficient; SRM, standardised response mean; WOMAC, Western Ontario and McMaster Universities osteoarthritis index

Table 1 Baseline characteristics of the patients

	Full WOMAC function subscale* (n = 1218)	Incomplete WOMAC function subscale† (n = 144)
Age (years)	66.9 (10.4)	69.7 (10.8)
Female sex	854 (70.1%)	59 (41.0%)
Body mass index (kg/m ²)	27.6 (4.7)	27.8 (5.0)
Disease duration (years)	4.5 (5.6)	4.7 (5.9)
Kellgren and Lawrence grade		
II	246 (20.2%)	19 (13.2%)
III	530 (43.6%)	62 (43.1%)
IV	440 (36.2%)	63 (43.8%)
NSAID intake during the past 4 weeks	355 (29.2%)	38 (26.4%)
Analgesics‡ during the past 4 weeks	699 (57.5%)	87 (60.4%)
Symptomatic slow acting drug intake§ during the past 4 weeks	437 (36.0%)	38 (26.4%)
Pain (0–100 VAS)		
Mean (SD)	57.9 (17.0)	57.6 (17.4)
Week 0 to week 4 (SD)	–23.3 (22.1)	–21.4 (22.4)
Global assessment (0–100 VAS)		
Mean (SD)	58.6 (19.0)	59.3 (20.6)
Week 0 to week 4 (SD)	–23.1 (24.1)	–21.0 (26.6)
WOMAC function score (0–68)		
Mean (SD)	29.7 (11.4)	–
Week 0 to week 4 (SD)	–7.8 (9.7)	–

Values are mean (SD) or n (%).

*Patients without missing data for the WOMAC function subscale at the baseline visit; thus those who were involved in the derivation process.

†Patients with missing data for the WOMAC function subscale or who did not complete the questions at all at baseline.

‡Other than NSAID.

§Chondroitin sulphate, diacerheine, or avocado/soybean unsaponifiables.

NSAID, non-steroidal anti-inflammatory drugs; VAS, visual analogue scale; WOMAC, Western Ontario and McMaster Universities osteoarthritis index.

To assess the test–retest reliability of the resulting WOMAC function subscale short form, a subsample of 86 patients was asked to complete the full WOMAC function subscale again, 48 hours after the baseline visit. These patients had begun taking NSAIDs 48 hours after the baseline visit (that is, after completing the WOMAC function subscale a second time).

Statistical analysis

First, we computed descriptive statistics on clinical and demographic variables. Then we used a four step procedure to eliminate items as follows:

- Step 1. We ranked the 17 items of the complete WOMAC function subscale from highest to lowest importance according to the patients' and rheumatologists' opinions, excluding the five items that were least important for both patients and rheumatologists. The whole sample was then divided into tertiles of the WOMAC function subscale score to investigate the potential impact of the level of functional impairment on the patients' ranking.
- Step 2. We ranked the 17 items by the proportion of missing data per item. Items with a high proportion of missing data were excluded.
- Step 3. Items whose distribution of answers showed a floor or ceiling response were excluded. This response is present when most of the answers are clustered in only a few response options at one extreme—that is, when most of the subjects attest to having no difficulty (floor response) or extreme difficulty (ceiling response) in the activity. For floor response items, it is impossible to detect improvement, while for ceiling response items, it is not possible to distinguish among various grades of difficulty, as most of the subjects answer the same way.
- Step 4. We tested for potentially redundant items. Inter-item correlation coefficients were computed. When the

correlation coefficient was greater than 0.75, the least important item of the pair in the patients' ranking was excluded.

Responsiveness was assessed by use of the standardised response mean (SRM) for the complete WOMAC function subscale and the WOMAC function short form. SRM is the mean change in score between the baseline and the final visit divided by the standard deviation of the change in score. Test–retest reliability was assessed using the intraclass correlation coefficient (ICC). Construct validity of the WOMAC function short form was assessed using the correlation between scores of the long and short forms, as recommended when the original scale cannot be considered a gold standard (that is, the reference measurement instrument).² Internal consistency was assessed using Cronbach's α .⁹

Statistical analyses involved use of the SAS Release 8.2 statistical software package.

RESULTS

In all, 1362 patients were enrolled in the study: 1019 (75%) with knee osteoarthritis and 343 (25%) with hip osteoarthritis. At the baseline visit, 1218 patients (89%) completed the full WOMAC function subscale without any missing data. The derivation process is based on these 1218 patients, described in table 1.

Ranking of the 17 items of the complete WOMAC function subscale

Patients and rheumatologists were consistent in ranking the importance of items (table 2). The four most important items for rheumatologists were among the five most important items for patients. The five least important items for rheumatologists were among the six least important items for patients. The ranking of item importance was similar

between patients with hip osteoarthritis and those with knee osteoarthritis (data not shown), except for “descending stairs” (ranked sixth and first, respectively), and “putting on socks/stockings” (first and 12th, respectively). As these items are relatively specific to the location of the osteoarthritis (hip or knee), this discrepancy was expected. The ranking of the items’ importance was similar between men and women (data not shown), except for “going shopping” (ranked 10th and fourth, respectively) and “performing light domestic duties” (ranked 13th and seventh, respectively).

The five least important items for both patients and experts were “lying in bed,” “bending to the floor,” “rising from bed,” “sitting,” “taking off socks/stockings,” and “standing.”

Results of dividing the whole sample into tertiles of the WOMAC function subscale score showed exactly the same items being selected by patients in the three subgroups.

Ranking of the 17 items by the proportion of missing data

Three items generated notably more missing data than the others. These items may have been interpreted too literally and considered not to be relevant—for example, domestic duties may have been interpreted only as cleaning the house and therefore probably of more concern to women, while respondents answering the getting in/out of the bath question may not have appreciated that this question can alternatively be considered relevant to getting in/out of the shower.

The items excluded were “performing heavy domestic duties,” “performing light domestic duties,” and “getting in/out of the bath.”

Items for which the distribution of answers showed a floor or ceiling response

Almost all the items of the complete WOMAC function subscale had a good distribution of answers among response modes. However, two had a saturation point in one or two response modes: for “bending to the floor” and “lying in bed,” 74% and 75% of the answers, respectively, were “no difficulty” or “slight difficulty.”

The items excluded were “bending to the floor” and “lying in bed.” Both items had already been excluded in a previous step.

Inter-item correlation

Pairs of highly correlated items ($r > 0.75$) were “putting on socks/stockings” with “taking off socks/stockings” ($r = 0.85$) and “performing light domestic duties” with “performing heavy domestic duties” ($r = 0.78$).

The items excluded were “taking off socks/stockings,” and “performing heavy domestic duties.” Both items had been excluded in a previous step.

Summary of the reduction procedure

The eight items of the WOMAC function subscale short form derived by the above mentioned methods are shown in the appendix. These items were the eight most important in the patients’ opinion.

When summarising the different steps in the reduction procedure, it can be seen that six of the nine excluded items were excluded in at least two steps (two steps for four of the items and three for two of the items).

Performance characteristics

The WOMAC function subscale short form did not differ substantially from the complete WOMAC function subscale either in responsiveness (SRM = 0.84 ($n = 1169$) and 0.80 ($n = 1048$), respectively) or in test–retest reliability (ICC = 0.75 (0.65 to 0.83) and 0.79 (0.69 to 0.87), respectively).

Construct validity of the WOMAC function subscale short form was excellent ($r = 0.95$ between the long and short forms). Internal consistency was good in the WOMAC function subscale short form and the complete WOMAC function subscale ($\alpha = 0.84$ and $\alpha = 0.93$, respectively).

DISCUSSION

Using patients’ and rheumatologists’ opinions in France, and based on the Likert version of the French Canadian WOMAC function subscale, we have specified a short form of this subscale for patients with hip or knee osteoarthritis (including a broad spectrum of disease severity). To address recent recommendations for shortening composite measurement scales,² we have ensured that the original scale was valid, relevant in the context of hip and knee osteoarthritis, and had satisfactory measurement properties.^{4, 5}

The WOMAC function subscale short form contains only eight items. It was derived by preserving face validity

Table 2 The 17 items of the complete WOMAC function subscale ranked in importance by patients and rheumatologists

Item	Patients’ opinions* (n = 1349)	Rheumatologists’ opinions† (n = 497)	Patients’ ranking‡	Rheumatologists’ rankings§	Proportion of missing data (%)
Descending stairs	64.10	73.85	1	1	0.22
Ascending stairs	62.65	47.90	2	5	0.29
Walking on the flat	50.20	28.15	3	8	0.15
Getting in/out of a car	41.45	51.90	4	4	0.22
Rising from sitting	39.90	66.00	5	3	0.37
Going shopping	34.10	19.30	6	10	0.95
Getting on/off the toilet	33.65	28.95	7	7	0.37
Putting on socks/stockings	30.85	66.80	8	2	0.44
Getting in/out of the bath	30.00	45.65	9	6	4.26
Performing light domestic duties	26.65	10.25	10	12	4.77
Performing heavy domestic duties	26.30	23.75	11	9	5.29
Standing	25.70	9.05	12	13	0.22
Taking off socks/stockings	11.95	17.10	13	11	0.59
Sitting	8.40	3.20	14	15	0.37
Rising from bed	7.05	5.65	15	14	0.37
Bending to floor	4.60	2.00	16	16	0.22
Lying in bed	2.40	0.40	17	17	0.22

Each patient selected the five items they considered the most important to be improved by treatment; each rheumatologist selected the five items for which their patients are generally most impaired. Excluded items are in italics.

*Percentage of patients who considered this item as one of the five most important.

†Percentage of rheumatologists who considered this item as one of the five most important.

‡Rank of the item based on the percentage of patients who considered this item as one of the five most important.

§Rank of the item based on the percentage of rheumatologists who considered this item as one of the five most important.

(patients' and rheumatologists' opinions) and quality of the items (few missing data, no redundancy, good distribution of the answers across response modes). Preserving face validity is important because it increases the acceptance of the instrument by those who will ultimately use it and thus decreases the amount of missing data.³ This short form has good responsiveness, good test–retest reliability, and good construct validity for this sample, but these parameters should be validated in an independent sample of subjects from the target population.¹⁰ Our reduction procedure involved deleting items that were highly correlated, and thus a lower internal consistency was expected for the short form than for the complete subscale (an internal Cronbach's $\alpha = 1$ indicates redundancy).

As the WOMAC subscale is dedicated to patients with hip or knee osteoarthritis, our sample reflects this target population well. The proportion of patients with hip and knee osteoarthritis (three quarters knee and one quarter hip) is close to the distribution in the community.¹¹ As shown in table 1, our sample, is similar to samples included in trials on osteoarthritis treatment and represents a large spectrum of disease severity. Inclusion criteria, especially the requirement for NSAID treatment, were the same as those in the validation study of the WOMAC scale by Bellamy and associates.⁴ In our sample, the same items were selected by patients across the range of osteoarthritis severity: the results of dividing the sample into tertiles of the WOMAC function subscale score showed that the five least important items to patients (those to be excluded) were exactly the same in the three tertiles.

It has been assumed that items for assessing knee osteoarthritis may be somewhat different from those required for hip disease. In fact, when we evaluated the ranking of the 17 items of the complete WOMAC function subscale according to their importance to patients with hip or knee disease, the five least important items (those to be excluded) were the same for patients with both types of osteoarthritis.

According to previous recommendations, when the original scale cannot be considered a gold standard (the reference measurement instrument), an expert based approach to item reduction may be preferable to a statistical approach.² This situation is far more likely in the patients' self assessment of symptoms. An expert based approach has been employed in very few studies that involved reducing indices, and mainly served to help choose among several solutions provided by statistical methods.² We chose the other route. We used an expert based reduction procedure, and statistical analyses of the quality of the items were secondary criteria. To reduce information bias in the reduction process, we combined two types of expert: patient experts, who had experience of the problems (representatives of the target population), and rheumatologist experts, using their knowledge of a broad spectrum of the disease.

The originality of our approach lies in the large number of experts involved in the study. Expert based approaches usually rely on the authors' own judgment of redundancy and insufficient face validity, or on the use of consensus methods with relatively small panels of experts. For instance, Guillemin and colleagues¹² used two panels when shortening the arthritis impact measurement scales 2 (AIMS2): one of 19 experts (rheumatologists, rehabilitation specialists, and methodologists) and another of 12 patients. Whitehouse *et al*¹³ used a panel of 36 experts (orthopaedic surgeons, rheumatologists, nurses, physiotherapists, and research personnel). The large sample of patient experts ($n = 1218$) and rheumatologist experts ($n = 399$, approximately 15% of the rheumatologists in France) in our study is a good indicator of its representativeness and of the validity of the results.

The relevance of our reduction procedure is reinforced by the outcome of the procedure. The remaining items are the eight most important in the patients' opinion, and most of the excluded items were excluded in at least two steps of the reduction procedure. Taking patients' opinion into account in deriving short forms of validated questionnaires could improve the clinical relevance of such methods.

Whitehouse *et al*¹³ proposed a seven item short form of the WOMAC function subscale, but the derivation process involved only a subgroup of patients with severe disease (patients undergoing hip or knee arthroplasty). In this context, the short form should be dedicated to assessing the outcome of total joint arthroplasty, as Whitehouse indicated. However, five items are shared between the Whitehouse form and our own. The particular population in Whitehouse's study may explain some of the discrepancies between the two short forms—especially that fact that activities such as "sitting" or "rising from bed," which are more likely to be impaired in severe disease, are two of the seven items included in Whitehouse's version but excluded from our version (because they were ranked 14th and 15th, respectively, by the patients).

The assessment of the performance characteristics of the WOMAC function subscale short form, its clinical relevance, and its acceptability require further studies in independent samples. Such studies should involve different versions of the WOMAC function subscale, as well as different language translations and different scaling formats, and should be conducted in different countries, in different clinical environments (for example, rheumatology, orthopaedic surgery, physiotherapy, rehabilitation), and with different interventions.

ACKNOWLEDGEMENTS

This study was supported by an unrestricted grant from Merck, Sharp & Dohme Chibret Laboratories, France.

Authors' affiliations

F Tubach, G Baron, P Ravaud, Département d'Epidémiologie, Biostatistique et Recherche Clinique, Groupe Hospitalier Bichat-Claude Bernard (Assistance Publique – Hôpitaux de Paris), Faculté Xavier Bichat (Université Paris 7), Paris, France

B Falissard, Faculté Paris XI, Département de santé publique, Hôpital Paul Brousse (Assistance Publique – Hôpitaux de Paris), Villejuif, France

I Logeart, Merck Sharp & Dohme, Chibret Laboratories, Paris, France

M Dougados, Service de Rhumatologie B, Hôpital Cochin (Assistance Publique – Hôpitaux de Paris), Faculté Paris V, Paris, France

N Bellamy, CONROD, Faculty of Health Sciences, The University of Queensland, Brisbane, Queensland, Australia

APPENDIX

Proposed WOMAC function subscale short form (eight items)

- Descending stairs
- Ascending stairs
- Rising from sitting
- Walking on flat
- Getting in/out of a car
- Going shopping
- Putting on socks/stockings
- Getting on/off the toilet

The WOMAC function subscale gradations in the Likert-scaled French Canadian 3.0 version are: 0 = none, 1 = slight, 2 = moderate, 3 = severe, 4 = extreme.

The WOMAC function subscale short form comprises a total of 32 possible points, with 0 being the best and 32 being the worst.

REFERENCES

- 1 **Moran LA**, Guyatt GH, Norman GR. Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *J Clin Epidemiol* 2001;**54**:571–9.
- 2 **Coste J**, Guillemin F, Pouchot J, Fermanian J. Methodological approaches to shortening composite measurement scales. *J Clin Epidemiol* 1997;**50**:247–52.
- 3 **Streiner D**, Norman G. *Health measurement scales. A practical guide to their development and use*, 2nd ed. Oxford: Oxford University Press, 1995.
- 4 **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.
- 5 **McConnell S**, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum* 2001;**45**:453–61.
- 6 **Altman R**, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al*. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;**34**:505–14.
- 7 **Altman R**, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, *et al*. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum* 1986;**29**:1039–49.
- 8 **Choquette D**, Bellamy N, Raynauld J. A French-Canadian version of the WOMAC osteoarthritis index. *Arthritis Rheum* 1994;**37**(S226).
- 9 **Cronbach LJ**. Coefficient alpha and the internal structure. *Psychometrika* 1951;**16**:297–334.
- 10 **Coste J**, Fermanian J, Venot A. Methodological and statistical problems in the construction of composite measurement scales: a survey of six medical and epidemiological journals. *Stat Med* 1995;**14**:331–45.
- 11 **Felson DT**, Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. *Arthritis Rheum* 1998;**41**:1343–55.
- 12 **Guillemin F**, Coste J, Pouchot J, Ghezail M, Bregeon C, Sany J. The AIMS2-SF: a short form of the Arthritis Impact Measurement Scales 2. French Quality of Life in Rheumatology Group. *Arthritis Rheum* 1997;**40**:1267–74.
- 13 **Whitehouse SL**, Lingard EA, Katz JN, Learmonth ID. Development and testing of a reduced WOMAC function scale. *J Bone Joint Surg Br* 2003;**85**:706–11.