Author copy:

Whitehouse SL, Learmonth ID, Crawford RW. Validation and treatment of missing values for the reduced WOMAC function scale. *J Orthop Surg (Hong Kong) - In Press February 2007*

Validation and Treatment of Missing Values for the Reduced WOMAC Function Scale

SL WHITEHOUSE PHD — Research Fellow, §

ID LEARMONTH FRCS — Professor of Orthopaedic Surgery, ¥

RW CRAWFORD DPHIL — Professor of Orthopaedic Research, §

§ Orthopaedic Research Group, Queensland University of Technology, Level 5, Clinical Sciences Building, Prince Charles Hospital, Chermside, QLD 4032, Australia

¥ Department of Orthopaedic Surgery, University of Bristol, Winford Unit, Avon Orthopaedic Centre, Bristol BS10 5NB UK

Corresponding author: Sarah Whitehouse

Tel: +61 (0)7 3350 8481

Fax: +61 (0)7 3350 8043

e-mail: s.whitehouse@qut.edu.au

Abstract

Aim*:* The reduced WOMAC function scale has been developed and initial validity performed. However, further validation and recommendations for the treatment of missing values are required

Method: A cross-over study of 100 pre-operative TKA patients was performed for further validation. Repeatability testing was performed for full and reduced WOMAC scores. Missing value protocols were developed, examining number of valid responses, means and SDs.

Results*: Cross-over:* There was no significant difference between full and reduced scales.

Test-retest: There was no significant difference between scores for either the full or reduced scales.

Missing values: The frequencies of valid responses, means and SDs when using different missing value protocols (none missing, 0-1, upto 2 and upto 3 missing), indicated that the 1 or 2 missing protocol was optimal.

Conclusions: The reduced WOMAC function scale has been further validated. It is proposed that where 1 or 2 items are missing, the average value for the sub-scale is substituted in lieu of these missing values.

Keywords: Reduced WOMAC, missing values, validation, outcome

Introduction and Aims

The Western Ontario and McMasters Universities Osteoarthritis Index (WOMAC) is a self assessed, disease-specific measure for patients with osteoarthritis of the hip and knee. The original version comprises 24 items in three dimensions: pain (five items), function (17 items) and stiffness (two items). (1), (2).

An abbreviated version of the WOMAC function scale has been developed and initial validity performed. (3) Initial validity was established by means of criterion validity, convergent construct validity, reliability and responsiveness using the dataset also used for the derivation of the scale. However, validation using independent datasets to further assess validity as well as repeatability, and recommendations for the treatment of missing values is required.

The aims of this paper are three-fold:

1) demonstrate the criterion validity of the reduced scale by means of a cross-over study,

   Criterion validity is the correlation of a scale with some other measure of the trait under study, ideally a 'gold standard' which has been used and accepted in the field. In the case where an existing scale has being modified (or reduced as is the case here) then comparison with the original, full scale is necessary.

2) assess repeatability of the reduced scale,

Test-retest reliability (repeatability) is the relationship between scores obtained by the same person on two or more separate occasions.

3) determine a missing values protocol for the reduced scale.

The standard protocol for missing values, as indicated in the WOMAC User Guide(2) is where, if there are 4 or more of the 17 function items missing, then the patient's response is regarded as invalid, and the sub-scale score is disregarded in further analysis. Where there are 1-3 items missing, then the average value for the sub-scale is substituted in lieu of these missing values.

No such protocol has been established for the reduced scale. In this paper, the protocol for the consideration of missing item responses is addressed.

All scores are given on a 0-100 scale, worst to best.

Patients and Methods

Cross-over study to further demonstrate validity

Between April and September 2000, 100 consecutive patients attending a pre-operative assessment clinic for consideration for total hip or total knee replacement at the UK institution were approached for enrolment in the randomised crossover validation study.

Consenting patients were randomised to receive either the full or reduced WOMAC function scale, along with the WOMAC pain dimension. Patients continuing on to be admitted were given the alternate version on admission (usually 2-3 weeks after attending pre-assessment clinic).

The combination of the original WOMAC pain dimension and the original, full function dimension is referred to as the ‘full scale’. Similarly, the original WOMAC pain dimension with the reduced WOMAC function scale is the ‘reduced scale’.

The study group comprised 66 patients of the consenting 100 who continued on to admission within the study period. Of the 34 patients who did not proceed to surgery, the operation was deferred for the majority due to problems detected in the pre-operative assessment visit. Twenty one of these patients completed the reduced scale, and 13 the full scale.

Of the 66 patients who underwent surgery, there were 36 total knee replacements and 30 total hip replacements. Twenty nine of these 66 (43.9%) received the full scale at pre-assessment clinic, and then the reduced scale on admission, 31 (47.0%) received the reduced scale at pre-assessment clinic and the full scale at admission, and the remaining 6 (9.1%) had both scales on the same day, as they were admitted

straight from pre-assessment clinic. In these cases, administrations of the questionnaires were hours apart.

For the study group of 66 patients, the median time between administrations of each questionnaire was 14 days (range 0 to 72 days).

Test-retest study to assess repeatability

Between May and August 2003, patients attending clinics (both pre- and post-operative) for total joint replacement at the Australian institution were invited to participate in a test retest study of either the full or reduced WOMAC scores (run concurrently).

Consenting patients were asked to complete a questionnaire, with a second one being posted out approximately 2 weeks later. Those pre-op patients being admitted before this date were asked to complete the questionnaire just before surgery.

Fifty five patients were recruited to complete the full score and 53 for the reduced scale. Of these, 51 successfully completed the two full scales (92.7%), with 52 (98.1%) completing both reduced scores. Those patients not completing the second questionnaire before their surgery were excluded from the study.

All of those patients completing the full scale were seen pre-operatively, two were for revision surgery. Twenty four (47.1%) were seen for hip

problems and 27 (52.9%) for their knees. The reduced cohort consisted of 41 (78.8%) pre-operative and 11 (21.2%) post-operative patients, with one revision patient. Nineteen (36.5%) were seen for their hip, 32 (61.5) for their knees and one (1.9%) for both.

For the full scale group, the median time between administrations was 13 days (range 1 to 36) and 7 days (range 2 to 19) for the reduced scale group.

Missing values to determine missing value protocol

For this section of analysis, data from the original WOMAC reduction paper (3) was utilised. The datasets were from the Kinemax Outcomes Study (KOS) (4), which is an international, prospective cohort study of primary total knee replacement (TKR) for patients with osteoarthritis, and the US Medicare Beneficiaries Hip Replacement Study that has been described fully elsewhere (3). Patients were recruited between September 1997 and December 1998 and all surgeons used the Kinemax Plus (Stryker Howmedica Osteonics) total knee replacement prosthesis. Data were gathered pre-operatively and at 3 and 12 months post-operatively by means of a physical examination (Knee Society Score) and a self-completed questionnaire booklet, which comprised of demographic and socio-economic details, as well as information regarding self-reported comorbid conditions, SF-36, WOMAC, patient expectation (pre-operatively), and at follow-up included 4 patient satisfaction questions which could be combined to give a summary satisfaction score from 0 – 100 (5). Data from the US Medicare

Beneficiaries study comprised 3 year postoperative WOMAC scores. The 3 and 12 month postoperative TKR and 3 year postoperative THR datasets were utilised. This is due to the fact that the pre-operative TKR data were screened for completeness as part of the inclusion criteria – that patients should be able to complete the questionnaires. Hence an exercise examining missing values is not relevant for this data.

The numbers of missing scores for the reduced scale were examined where all responses were present, or where 1, 2 or 3 responses were missing. Calculations for more than 3 responses missing were not made, as these would not be valid for the full scale. Also, average score imputation is usually restricted to cases where at least half of the items in the scale are completed. (6) Where missing values were present, the average value for the scale was substituted, as recommended for the full scale. The numbers of missing values observed were compared to the valid number of scores for the full scale.

Results

Cross-over study

Initially, pain scores were compared for each administration in the study group of 66 patients. All patients had a valid pain score calculated. Comparisons were made using the paired t-test, as the distribution of the differences between scores is assumed to be Normal. These

values are given in Table I. There was no significant difference between these scores at the 5% level. All p-values are two tailed.

Similarly, scores for the function scales were compared. These results are also given in Table I. Again, there was no significant difference between the full and reduced scales at the 5% level. For both scales, the missing item protocols discussed previously were implemented. There are only 65 valid scores for comparison in this group as 1 response for the full scale had only 8 items completed and hence a valid score could not be derived. It is interesting to note that for the reduced scale, only 1 item (item 14 on original scale – sitting) was missing for the whole study group, whereas missing items for the full scale are represented in Figure I.

The frequencies of missing items for the full WOMAC function scale are given in Figure I. In all but one instance, where missing items have occurred in the full scale, the same patient has completed all items in the reduced scale, even if this involves one of the items they previously missed (there are items in the reduced scale that were missing in the full version).

Test-retest study

The mean scores for the full and reduced scales at each administration are shown in Table II. As data was non-parametric in nature, non parametric tests are utilised (Wilcoxon signed rank test and Spearman's correlation coefficient). The correlations between the scores were

calculated and are also presented in Table II. These were all significant at the 5% level indicating excellent agreement in the scores when repeated.

When comparing the scores using the Wilcoxon signed rank test for paired data, there was no significant difference at the 5% level for each of the dimensions in each scale.

Missing values

The numbers of missing values observed were compared to the valid number of scores for the full scale. These are represented in Table III.

For both the 12 month postoperative TKR and the THR datasets, there are more valid responses with up to 3 items missing for the reduced scale than for the full scale. This is due to the fact that those missing items invalidating the full scale are not necessarily those which have been retained in the reduced scale, that is that those missing from the full scale may be those which were not included in the reduced scale (perhaps for this reason). This means that a valid score can be calculated for the reduced scale, but not for the full scale. It is salient that those items with high frequencies of missing responses were purposefully removed from the scale, and so the number of missing values for these items should be reduced. It should also be considered that the fact that the number of items in the scale has been reduced, also means that the influence of each item on the scale score is

proportionately increased (from 6% per item to 14%), and hence the impact of missing items is larger for the reduced scale.

There does not appear to be a substantial increase in numbers of valid responses between the ‘up to 2 missing’ and ‘up to 3 missing’ response protocols, particularly when the number of valid full scale responses is considered. For two if the three datasets, the number of valid scores is at least or more than those for the full scale. For this small gain, the supposition that the completed items are representative of the missing ones rises from 29% (2 of 7 items) to 43% (3 of 7 items). Although in general this may be deemed acceptable in terms of score calculation (6), in this case it was decided that there was not a sufficient gain in the number of valid responses to warrant this increase in supposition.

It is therefore proposed that the standard protocol for the treatment of missing values for the reduced scale is that where there are 3 or more of the 7 function items missing the patient’s response is regarded as invalid, and the sub-scale score is disregarded in further analysis. Where there are 1 or 2 items missing, the average value for the sub-scale is substituted in lieu of these missing values.

The validity of this protocol was confirmed by comparing the means and standard deviations of both the TKR and THR data, incorporating those responses which had previously been considered invalid for the reduced scale, with both the full scale scores and those obtained when

all items in the reduced scale are present. These values are given in Table IV.

Discussion

Cross-over study

This randomised crossover study illustrates that there is no statistically significant difference between the reduced WOMAC function scale and the full WOMAC function scale when administered separately (all analyses previously have been on data gathered as the full scale and the reduced scale calculated from it as a subset).

It also highlights the fact that there is an increase in compliance when asking the respondent to complete an appreciably smaller set of questions. Patients missing items in the full scale have completed the same item when confronted with them in the reduced scale. This has important implications for data collection, completion and analysis.

Test-retest study

The test retest study for both the full and reduced scales indicates that both scales have high repeatability and that this important reliability measure is not diminished with the reduction of the scale.

Missing values

The results from this study indicate that the missing value protocol implemented is not only valid, as seen in Table IV, but will also add power to the results of some studies, particularly where relatively small numbers are involved, by ensuring that the data eligible for analysis is maximised.

Conclusion

In conclusion, this reduced version of the function dimension of the WOMAC retains excellent validity, reliability and responsiveness. Its use is recommended along with the original pain dimension in studies of total joint replacement.

Allowance for missing values can be made where, if there are 3 or more of the 7 function items missing, the patient's response is regarded as invalid, and the sub-scale score is disregarded in further analysis. Where there are 1 or 2 items missing, then the average value for the sub-scale is substituted in lieu of these missing values.

The reduced WOMAC function scale provides an alternative which will limit the number of missing values, and hence the number of invalid WOMAC scores. This has important implications when considering power studies for smaller studies, and the impact of questionnaire choice on compliance and duplication of data, as well as data completion, collection and analysis.

Tables:

Dimension	Full scale (SD)	Reduced scale (SD)	Number of subjects	Paired t-test statistic *(p-value)*
Pain	61.7 (17.8)	60.8 (19.1)	66	0.58 *(p = 0.56)*
Function	60.5 (16.7)	59.9 (17.7)	65	0.46 *(p = 0.65)*

Table I. Means and statistical significance tests of each group

Scale	Administration 1	Administration 2	Correlation
Full pain	46.3 (17.1)	47.6 (16.7)	0.77 (p<0.001)
Full function	42.2 (19.3)	41.6 (19.0)	0.79 (p<0.001)
Reduced pain	51.4 (22.0)	51.1 (20.2)	0.86 (p<0.001)
Reduced function	49.5 (22.8)	47.5 (22.2)	0.83 (p<0.001)

Table II. Mean (standard deviation) of each scale and Spearmans rho (p-value).

Criteria	3 month TKR	12 month TKR	3 year THR
Full ≤ 3 missing	806	762	841
Reduced 0 missing	766	708	790
Reduced 0 or 1 missing	801	756	834
Reduced 0, 1 or 2 missing	804	763	841
Reduced 0, 1, 2 or 3 missing	806	766	849

Table III. Valid score frequencies for various protocols for missing responses

Criteria	3 month TKR	12 month TKR	3 year THR
Full ≤ 3 missing (SD)	70.0 (18.8)	73.8 (20.6)	78.6 (18.9)
Reduced 0 missing (SD)	70.6 (18.4)	75.9 (18.4)	78.6 (18.4)
Reduced 0, 1, 2 or 3 missing (SD)	70.7 (18.5)	75.3 (20.5)	78.4 (18.5)

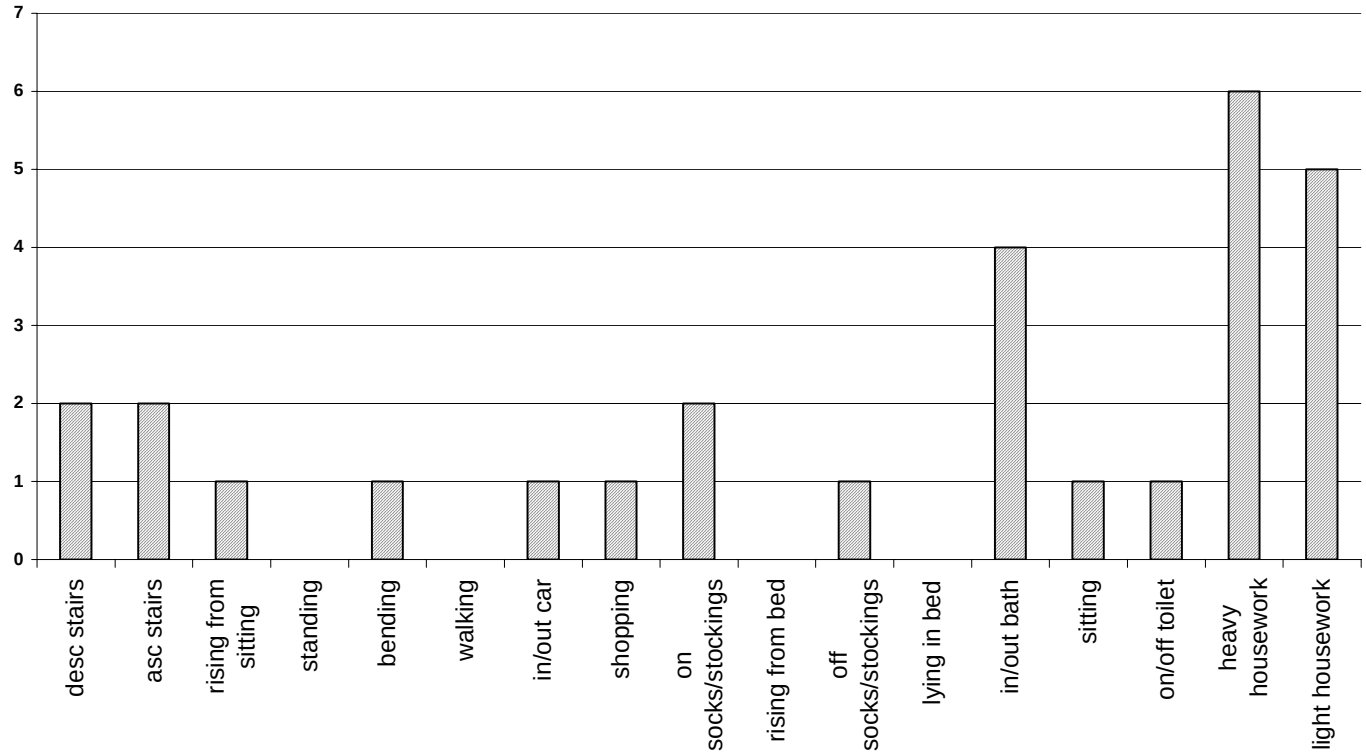
Table IV. Mean scores and standard deviations using different criteria for missing values

List of legends for figures:

Figure I. Missing item frequencies for full WOMAC function scale

Figures:

Figure I. Missing item frequencies for full WOMAC function scale



References

1. Bellamy N, Buchanan W, Goldsmith C, Campbell J, Stitt L. Validation study of the WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthopaedic Rheumatology 1988;1:95-108.

2. Bellamy N. WOMAC osteoarthritis index: user guide IV. 2000.

3. Whitehouse SL, Lingard EA, Katz JN, Learmonth ID. Development and testing of a reduced WOMAC function scale. J Bone Joint Surg Br 2003;85(5):706-11.

4. Lingard EA, Katz JN, Wright E, Sledge CB, Group KO. Predicting the outcome of total knee arthroplasty. J Bone Joint Surg [Am] 2004;86-A(10):2179-86.

5. Mahomed NN, Sledge CB, Daltroy L, Fossel AH. Self-administered patient satisfaction scale for joint replacement arthroplasty. J Bone Joint Surg [Br] 1998;80(Suppl):9.

6. Staquet MJ, Hays RD, Fayers PM. Quality of life assessment in clinical trials. Methods and practice: Oxford Medical Publications; 1998.