COVER SHEET

Development and Testing of a Reduced WOMAC Function Scale

SL Whitehouse PhD
Research Fellow
Orthopaedic Research Unit
Queensland University of Technology
Prince Charles Hospital
Chermside, Brisbane QLD 4001
Australia

EA Lingard MPhil
Research Coordinator
Department of Orthopaedic Surgery
Freeman Hospital
Newcastle upon Tyne
NE7 7DN

JN Katz MD
Consultant Rheumatologist
Robert Brigham Arthritis & Musculoskeletal Diseases
Research Center
Brigham and Women's Hospital
Boston, USA

ID Learmonth FRCS
Professor in Orthopaedic Surgery
Department of Orthopaedic Surgery
University of Bristol
Winford Unit
Avon Orthopaedic Centre
Bristol BS10 5NB

Corresponding Author: ID Learmonth
Tel: *0117 928 2658*
Fax: *0117 929 4217*
e-mail: ian.learmonth@bristol.ac.uk

ABSTRACT

Objectives: To derive and validate a reduced WOMAC function scale.

Methods: We used prospective data from 862 primary TKR and 716 3-year post-op primary THR patients. The reduced scale was derived using input of clinical experts as well as analysis of data on patients undergoing TKR and THR. The scale was tested for validity, reliability and responsiveness.

Results: Items retained were: **ascending stairs, rising from sitting, walking on flat, getting in/out of car, putting on socks, rising from bed, sitting**.

The reduced and full scales had comparable, moderate correlations with other measures of function, confirming convergent validity. Cronbach's alpha was high (α>0.85) with the reduced scale, confirming reliability. Responsiveness (measured by standardised response means, calculated using change in status at 12-months from pre-op) was greater for the reduced scale (full=1.4, reduced=1.6).

Conclusions: This reduced version of the WOMAC function scale provides a practical, valid, reliable and responsive alternative to the full function scale for use following total joint replacement. However, further work is needed to demonstrate its wider applicability.

Key words: WOMAC, reduced WOMAC, function scale, TKR, THR, osteoarthritis, validate, reduction.

INTRODUCTION

Many instruments presently exist for patient and surgeon assessment of outcome after total joint replacement. These include generic (measures of general health status), disease specific and clinical evaluation measures.

The Western Ontario and McMasters Universities Osteoarthritis Index (WOMAC)[1,2] is a self-assessed, disease-specific measure for patients with osteoarthritis of the hip and knee, comprising 24 items in 3 dimensions; pain (5 items), function (17 items) and stiffness (2 items). There are 2 versions available. One has a visual analogue response scale, and the other a Likert 5 point response scale.

The Short-Form-36 (SF-36)[3] measures general health status with 36 items in 8 dimensions, 4 pertaining to mental health and 4 to physical health.

Investigations in the US[4] recommended that both disease-specific (e.g. WOMAC) and generic (e.g. SF-36) tools should be incorporated as standard when conducting clinical trials of outcome following total joint replacement. Inclusion of both instruments, often along with basic demographic data, makes data collection and analysis cumbersome and expensive, may reduce compliance, and also introduces some duplication of data.

An abbreviated 12-item version of the SF-36 (SF-12) has been developed and validated as a summarised generic health score[5,6,7]. Several studies[8,9] have shown that there is redundancy within the WOMAC function scale and have suggested that developing the scale further and omitting redundant items may be worthwhile. Reliability studies have also shown that the scale has values of Cronbach's alpha of greater than 0.9, indicating that the opportunity for trimming exists.

It is critical that this reduced score is representative of the full scale, relevant, user-friendly (both for ease of completion and time dependency), valid, reliable, and responsive.

As such, it is imperative that any new or improved scoring system is rigorously validated. As well as being representative of the full scale, the reduced WOMAC would also require evaluation in three key areas[10]:

- *Validity*: does it measure what it's supposed to measure? There are many types of validity - most important in this instance are criterion and convergent construct validity. The traditional definition of 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 this case, that means comparing the reduced with the full scale. However, as the reduced scale is essentially a component of the full scale, high correlations are expected and are used simply to ensure that spurious results are not confounded. Convergent construct validity examines how

strongly a new scale is associated with other measures of the same construct.

- *Reliability*: this term encompasses the internal consistency of a scale, usually given as Cronbach's alpha (α), which measures the degree of correlation amongst items. Values of Cronbach's alpha of greater than 0.7 indicate adequate reliability for a scale, whereas values above 0.9 may indicate redundancies in the scale[11]. (While Cronbach's alpha's greater than 0.9 are necessary for reliable individual scores, they are redundant for group means, the usual focus of research). Reliability also includes both between observer and test-retest reproducibility. As the WOMAC is a self-administered tool, between observer reproducibility is not relevant in this case, and as the new scale is a subset of the original, test-retest reliability is assumed to be inherent and has not been investigated further here.

- *Responsiveness*: is the measure able to detect true changes in clinical states when they occur? There are several ways of assessing responsiveness, but in this instance, the method of standardised response means (SRMs) has been utilised[12].

PATIENTS AND METHODS

The Kinemax Outcomes Study (KOS) is a multi-national, prospective cohort study of primary total knee replacement (TKR) for patient with osteoarthritis. Patients were recruited between September 1997 and December 1998 and all surgeons used the Kinemax Plus (Stryker Howmedica Osteonics) prosthesis. Data were gathered pre-operatively and at 3 and 12 months post-operatively with a physical examination and a self-completed questionnaire booklet.

Trained research assistants collected clinical history and physical examination data and the patient questionnaire included specific questions on function including walking distance, use of walking aid and stair climbing ability. From these data the Knee Society clinical rating system was used to derive a Knee Score and a Function Score[13].

Questionnaire books contained socio-economic data, self-reported comorbid conditions, WOMAC, SF-36 and at follow-up included 4 satisfaction questions that are combined to give a summary satisfaction score from 0-100[14]. On a subset of patients in the UK, the 12-item Oxford Knee Score[15] was also collected.

All data for unilateral (within 12 months) TKR patients from centres in the UK (Edinburgh, Newcastle, Mansfield, Derby Nottingham and Bristol), US (Boston, New York, Cedar Rapids and Corvallis) and Australia (Gold Coast and Adelaide) were included for analysis.

For purposes of further validation of the scale, we also utilised 3 year follow-up data on a random sample of 716 Medicare patients in the US who had primary

total hip replacement (THR) for osteoarthritis in 1995. These data included 3 year postoperative WOMAC, SF-12 and Harris Hip Score[16]

In order to reduce the WOMAC score successfully, several factors were taken into consideration. The stiffness score of the WOMAC is largely redundant and is commonly excluded from the questionnaire. The Pain Scale has just five items and no reduction was deemed necessary. Therefore item reduction was targeted to the 17-item function scale.

The item reduction was initiated by a clinically-driven process. A poll of orthopaedic and rheumatology personnel in the UK and US was conducted. Thirty-six members of the orthopaedic and rheumatological community responded, including consultant orthopaedic surgeons and trainees, consultant rheumatologists, nursing staff, physiotherapists and research personnel, both in the UK and the US. Respondents were requested to indicate which 5 items from the function scale they would keep, using the following three criteria: (1) most likely to change after surgery, (2) what patients care about the most, and (3) representative of a broad spectrum of activity levels. (one orthopaedic surgeon gave 6 preferences instead of 5).

Items that were considered to have gender or cultural ambiguity or to be open to misinterpretation were omitted from the reduced scale.

Data analysis was then performed in order to confirm that the selected items represented a range of difficulties, were clinically sensitive to detecting change, had few missing values and were applicable to both hip and knee patients. The mean item responses were examined to identify the difficulty of each question for the patient carrying out the activity pre-operatively and at 3-months post operatively. Although the data under examination in this instance is categorical and hence the median values should ideally be examined, this measure of central tendency is not precise enough for this purpose.

For the item analyses, responses were coded in the standard fashion for WOMAC with: 0 being ‘no difficulty’ and 4 being ‘extreme difficulty’ performing the physical function (intermediate answers being 1 – mild, 2 – moderate, 3 – severe)[2].

The derived reduced questionnaire was then assessed for validity, responsiveness and reliability using the 1-year follow up data from the TKR dataset. For these analyses summary scores for each patient were calculated for both the full and reduced function scales by taking the mean of all responses, multiplying by 25, then subtracting from 100. This transformed the raw WOMAC function score to a 0-100 scale, worst to best. A score of 0 indicates extreme restriction in all activities, and a score of 100 indicates no restriction for any item. If 4 or more of the function items were missing for the full scale a score was not calculated, as per the standard protocol for the handling of missing responses.

In order to verify criterion validity, the correlations between the full and reduced WOMAC function scores for the preoperative, 3-month and 12-month postoperative data were examined. In order to determine if there was differential reporting for these factors, subgroups by age, sex and country were also compared. At all stages, due to the nature of the scoring system (producing categorical data), the non-parametric Spearman's rank correlation coefficient is utilised rather than the parametric Pearson's correlation coefficient.

Convergent validity was assessed by comparing the strength of correlation of the full and reduced scores with several scales measuring similar outcomes. For the TKR patients these included the SF-36 physical component score and physical function score, the Knee Society function score and the Oxford Knee Score. For the THR patients these included the SF-12 Physical Component Score and the Harris Hip Score.

The statistical significance of correlations was compared using the Fisher's test of equality of two correlations for the same sample[17]. All p-values are 2 tailed.

Responsiveness was assessed using standardised response means (SRM's) calculated as the mean change in score from pre-operative to 12 months divided by the standard deviation of the change in score[12]. This parameter assesses the extent of improvement. Therefore, patients indicating that their quality of life (taken from question 2 in the SF-36) was the same or worse since their operation were excluded from this section of analysis. Given that total knee replacement has dramatic effects on pain and function, values of greater than 1 were indicative of adequate responsiveness[12]. In order to evaluate whether the change is relevant to the patient, responsiveness was also assessed by determining whether changes in the full and reduced scores correlated with other indicators of change in patients' clinical status including the patient's perceived perception of change in quality of life, general health and satisfaction with outcome. Higher correlations indicate greater responsiveness[18].

RESULTS

Data Sources

1) Clinical opinion of orthopaedic personnel.

The respondents comprised 21 surgeons, 6 research personnel, 5 orthopaedic nurses and 4 physiotherapists, of which 24 were from the UK, and 12 from the US.

2) TKR data.

This dataset contained 862 primary TKR patients. A total of 806 (94%) of these had valid WOMAC scores at their 3 month review, with 762 (88%) at 12 months. The mean age was 70 years, range of 38 to 90, and 59% were female. The majority of patients (50%) were from the UK, 31% the US and 19% Australia.

3) THR data.

There were 716 patients with a diagnosis of osteoarthritis included in the analysis. The mean age of this group of patients is 73.6 years, range 65 to 93 years, with a standard deviation of 5.5 years and there were 665 valid WOMAC function scores for this dataset.

A) Derivation of scale

i) Clinical opinion of orthopaedic personnel:

The results of this survey are given in Table I. Those items which were eventually selected for retention in the final reduced model, are indicated in bold type. As indicated, 5 of the top 7 items were kept in the reduced scale. Descending stairs was not included as one stair item was already in the scale, and the ascending stairs item proved to be more responsive to change over time. In order to avoid items that would not apply to one gender or cultural group, domestic duties, shopping and bathing items were not included. The patient-based data were then examined to justify the inclusion of these 5 and to provide the other 2 items. The 2 items lower down the ranking were used to create a broad range of item difficulty. These items were also considered basic activities of daily living that all patients would encounter every day.

ii) Data driven analysis using TKR data:

The analysis of the TKR cohort gave the mean item values indicated in Table II. For the preoperative data the mean score of the easiest question is 1.5, 1 being mild restriction, and 2 being moderate restriction of activity. The most difficult item has a mean score of 2.7, where 3 is severe restriction of activity.

Similarly for the 3-month postoperative data, mean item scores are presented. There is a shift down the scale, as expected after surgery, from moderate to no difficulty with activity, with the easiest item obtaining a mean score of 0.7, and the most difficult of around 1.8.

In addition, the numbers of missing values for each item were taken into account when considering retention of each item. These missing response frequencies are also shown in Table II. As shown, the number of missing responses for items 8 (going shopping), 13 (getting in/out of bath), 16 (heavy housework) and 17 (light housework) are particularly high.

The items in the reduced scale were: ascending stairs (Q2), rising from sitting (Q3), walking on flat (Q6), getting in/out of car (Q7), putting on socks (Q9), rising from bed (Q10), sitting (Q14).

B) Validation of scale using TKR and THR data:

The reduced scale was assessed with a variety of methods to examine validity, reliability and responsiveness.

i) Criterion Validity:

Spearman's correlation coefficient between the two scales was 0.96 for the knee dataset and 0.97 for the hip dataset. At all time points for both TKR and THR patients, comparisons of the full and reduced WOMAC function scores gave remarkably similar mean values and standard deviations (Table III). The very strong correlation between the two scales and high agreement in scores support the hypothesis that the reduced scale captures functional status as well as the original version.

Mean scores were calculated, stratifying the data by sex, age (by quartiles) and country to indicate if there was differential reporting for these factors. The findings indicate similar mean scores for the reduced and full scales, irrespective of assessment time and subdivision. In fact, the mean scores never differ by more than 2 points. The results obtained when stratifying by sex are presented graphically in Figure I. Similar findings emerge when stratifying by country and age quartiles.

ii) Convergent Construct Validity

Convergent construct validity was assessed by determining whether the reduced scale has similar strength of correlation with other scales. For this section of analysis, all pre-op TKR data were utilised, as well as the THR data to compare the WOMAC scale with the Harris Hip Scores and the SF-12 physical component scores. For the knee data, both full and reduced scales were correlated with SF-36 physical component score, SF-36 physical function score, Knee Society function score and Oxford knee score. These are shown in Table IV. All Spearman rank correlation coefficients are significant at the 1% level, and for the reduced scale, are an average of a mere 0.035 less than those for the full scale (95% confidence interval for the differences 0.025 to 0.045), supporting the hypothesis that the reduced scale is valid. Furthermore, for all the scales compared, for both the hip and knee, the correlations between the full and reduced scales and related measures did not differ at the 5% level, using Fisher's test of equality of two correlations for the same sample.

iii) Reliability

Internal consistency for the 2 scales was measured using Cronbach's alpha. The values shown for the full scale are extremely high, as shown in Table III, ranging from 0.95 to 0.97, while the values obtained for the reduced scale are slighter lower (0.87 to 0.93). Thus the reduced scale maintained excellent internal consistency.

At 3 months postoperatively, the SRM for both the full and reduced scales were 1.3. At 12 months these increased to 1.4 for the full scale, and 1.6 for the

reduced scale. These indicate adequate responsiveness for both scales, and that the reduced scale was at least as responsive as the full scale.

Spearman's non-parametric rank correlation coefficients were calculated to assess the association between the changes in scale scores and the patients perceived change in functional status; these are given in Table V. Similar values were obtained for both the full and reduced scales, and all are significant at the 1% level, supporting the responsiveness of the measures. The correlations of the full and reduced WOMAC changes to the other questions reflecting perceived changes and satisfaction with outcome were not significantly different at the 5% level.

DISCUSSION

Both psychometric and clinical approaches were used to develop a shortened version of the WOMAC function scale. The reduced scale compares favourably with the full scale overall.

Convergent validity was demonstrated by moderately strong correlation between various physical function scales for both the TKR and THR patients. These included the SF-36 physical component score, the SF-36 physical function score, the Knee Society function score, and the Oxford knee score for knees, and the Harris Hip Score and SF-12 physical component score for the hips.

A Cronbach's alpha reliability coefficient of greater than 0.7 is generally required for group comparisons. The values obtained in this study, although marginally less than those for the full scale, remain more than adequate whilst also eradicating redundancy. Indeed, these results further support the work of Ryser et al[9] which indicated that some redundancy occurred in the scale and there was scope for reduction by omission of redundant items.

SRM's obtained in this study indicate that the two scales have similar responsiveness, reinforcing the concept that the reduced scale adequately represents the full scale. Indeed SRM's for the reduced scale are slightly higher than those for the full score, which may indicate that it is slightly more responsive. However, further studies are required to support this.

The reduced scale also correlates significantly with various other measures of perceived functional change and satisfaction with outcome. This further indicates that the scale is responsive to meaningful changes, as observed by the patient.

As the reduced scale is a subset of the full scale, it will be relatively simple to compare results across studies using either form, especially as the WOMAC is the recommended disease-specific outcome measure. This will increase its acceptability and usefulness within the orthopaedic community. Other shortened measures of outcome exist, for example the Oxford Hip and Knee Scores and the Bristol Knee Score, but all have their disadvantages. The

Oxford scores, although brief and simple, may not be sufficiently specific for use after total joint replacement[19,20] and there are concerns regarding missing values[20]. The Bristol knee score has not been validated, and is used almost exclusively by the Bristol Knee Group.

Further work should be done in order to further validate this reduced scale, in particular for non-operative patients, patients undergoing THR and revision total joint arthroplasty. Test-retest reliability needs to be established for the reduced scale. Also, further studies investigating compliance and missing values are needed.

CONCLUSION

We have developed a shortened version of the WOMAC function scale. It retains excellent reliability, validity and responsiveness. We recommend use of the scale along with the original pain dimension in studies of total joint replacement.

Although this work presents a suggested subset of items for the reduced WOMAC function scale, its development and validation has been thus far limited to the orthopaedic field. It is acknowledged that further work is required to validate its use in non-surgical patients, to ensure broader acceptability, and this is to be strongly recommended.

Acknowledgements

We thank the Kinemax Outcomes Group; in particular the Research Assistants involved in data collection for their tireless work. We would also like to acknowledge the work of Dr Elizabeth Wright who provided computer programming support for these analyses.

The Kinemax Outcomes Group consists of the following members:

Princess Margaret Rose Hospital, Edinburgh, UK
William Gilespie, Colin Howie, Ian Annan, Alasitair Gibson, Judith Lane

Freeman Hospital, Newcastle Upon Tyne, UK
Ian Pinder, David Weir, Nigel Brewster, Karen Bettison

King's Mill Clinic, Mansfield, UK
Maurice Needoff, Roz Jackson,

Derbyshire Royal Infirmary, Derby, UK
Tim Wilton, Peter Howard

Queen's Medical Centre, Nottingham, UK
Ian Forster, Paul Szyprt, Chris Moran, David Whittaker, Mike Bullock, Zena Hinchcliffe

Avon Orthopaedic Centre, Bristol, UK
Ian Learmonth, John Newman, Chris Ackroyd, George Langkamer, Robert Spencer, Mark Shannon, Evert Smith, John Dixon, Sarah Whitehouse

Brigham and Women's Hospital, Boston, US
Clement Sledge, Frederick Ewald, Robert Poss, John Wright, Scott Martin, John Kwon, Yvette Valderrama

Beth Israel Medical Center, New York, US
Steven Harwin, Michael Lichardi

Iowa Medical Clinic, Cedar Rapids, US
Mark Mehlhoff, Linda Weiler, Tom Cahalan

Neuromuscular and Joint Center, Corvallis, CAN
Richard Cronk, Allyson Sandago

The Knee Centre, Gold Coast, AUS
Stephen Rackemann, Emma McLaughlin

Queen Elizabeth Hospital, Adelaide, AUS
Peter Lewis, Robert Bauze, Tom Stevenson, Gordon Morrison, Jane Clasohm

This study, along with the Kinemax Outcomes Study, has been supported by Stryker Howmedica Osteonics, New Jersey, USA and Limerick, Ireland.

The Medicare THR outcome study was funded by NIH grant P60 AR 47782.

Tables

Item	Number including item	Rank *most to least popular*
3. rising from sitting	**27**	**1**
6. walking on flat	**27**	**1**
2. ascending stairs	**20**	**3**
1. descending stairs	16	4
17. light domestic duties	14	5
7. getting in/out of car	**13**	**6**
9. putting on socks	**12**	**7**
13. getting in/out of bath	11	8
8. going shopping	8	9
4. standing	7	10
15. getting on/off toilet	6	11
5. bending to floor	5	12
10. rising from bed	**5**	**12**
12. lying in bed	4	14
11. taking off socks	2	15
14. sitting	**2**	**15**
16. heavy domestic duties	2	15

Table I. Number of clinical advisors (n=36) indicating items for retention in reduced scale (retained items in bold)

Item	Clinical		Pre-op			3 month		
	n	**Rank** most to least popular	**mean**	**Rank** hardest to easiest	**Missing responses** number	**mean**	**Rank** hardest to easiest	**Missing responses** number
1. descending stairs	16	4	2.72	1	9	1.52	4	12
2. ascending stairs	**20**	**3**	**2.66**	**3**	**11**	**1.36**	**7**	**19**
3. rising from sitting	**27**	**1**	**2.40**	**7**	**2**	**1.33**	**8**	**1**
4. standing	7	10	2.19	10	8	0.99	12	10
5. bending to floor	5	12	2.39	8	12	1.37	5	20
6. walking on flat	**27**	**1**	**2.25**	**9**	**15**	**0.94**	**13**	**8**
7. getting in/out of car	**13**	**6**	**2.47**	**6**	**9**	**1.61**	**2**	**6**
8. going shopping	8	9	2.60	4	22	1.36	6	29
9. putting on socks	**12**	**7**	**1.98**	**12**	**9**	**1.23**	**9**	**3**
10. rising from bed	**5**	**12**	**2.01**	**11**	**8**	**1.04**	**11**	**10**
11. taking off socks	2	15	1.88	13	9	1.11	10	7
12. lying in bed	4	14	1.62	15	9	0.85	16	6
13. getting in/out of bath	11	8	2.52	5	80	1.55	3	97
14. sitting	**2**	**15**	**1.54**	**17**	**9**	**0.73**	**17**	**7**
15. getting on/off toilet	6	11	1.59	16	7	0.87	14	1
16. heavy domestic duties	2	15	2.68	2	42	1.79	1	77
17. light domestic duties	14	5	1.71	14	35	0.87	15	20

Table II. Results of clinical advisors survey and the data-driven results including: mean item response, difficulty rankings and number of missing responses per item (pre-operative and 3 months postoperative TKR data). Retained items in bold.

Joint	Time	Full WOMAC	Reduced WOMAC	Full WOMAC	Reduced WOMAC
		Mean (95% CI))	Mean (95% CI)	Cronbach's α	Cronbach's α
Knee	Pre-op	45.3 (44.1 - 46.5)	45.6 (44.3 - 46.8)	0.95	0.87
	3 months	70.0 (68.7 - 71.3)	70.6 (69.3 - 71.9)	0.96	0.91
	12 months	74.1 (72.6 - 75.5)	76.0 (74.5 - 77.5)	0.97	0.93
Hip	3 year	79.9 (72.6 - 75.5)	78.4 (77.0 - 79.8)	0.96	0.90

Table III. Full and reduced scale means (range 0-100, 100 best), 95% confidence intervals and values of Cronbach's alpha in patients undergoing total hip and knee replacement.

		Full WOMAC	*Reduced WOMAC*	*n*	*p-value*
Knee	Full WOMAC	-	**0.96**		
	SF-36 physical components score	0.44	0.42	862	1
	SF-36 physical function	0.56	0.52	862	1
	Knee Society function score	0.47	0.45	862	1
	Oxford knee score	0.77	0.72	201	0.83
Hip	Full WOMAC	-	**0.97**		
	Harris Hip Score	0.73	0.69	665	0.79
	SF-12 physical component score	0.72	0.68	665	0.99

Table IV. Spearman's correlation coefficients between full and reduced WOMAC scales and various measures of functional states for pre-operative TKR and 3 year postoperative THR patients. All values indicate significant correlations at the 1% level. P-values indicated are from Fishers test of equality of two correlations.

	Δfull WOMAC	***Δreduced WOMAC***	***n***	***p-value***
Health compared to 1 year ago	0.37	0.36	762	0.91
Change in quality of life since TKR	0.38	0.42	762	1
Satisfaction score	0.30	0.35	762	1

Table V. Spearman correlation coefficients for the associations between perceived change in health status and quality of life and satisfaction with TKR outcome and change (Δ) in WOMAC function (calculated as the change in scores from pre-op to 12 months). All correlations are significant at the 1% level. The difference in correlations between full and reduced scores and each outcome are not statistically significant at the 5% level (p-values are indicated).

Figures

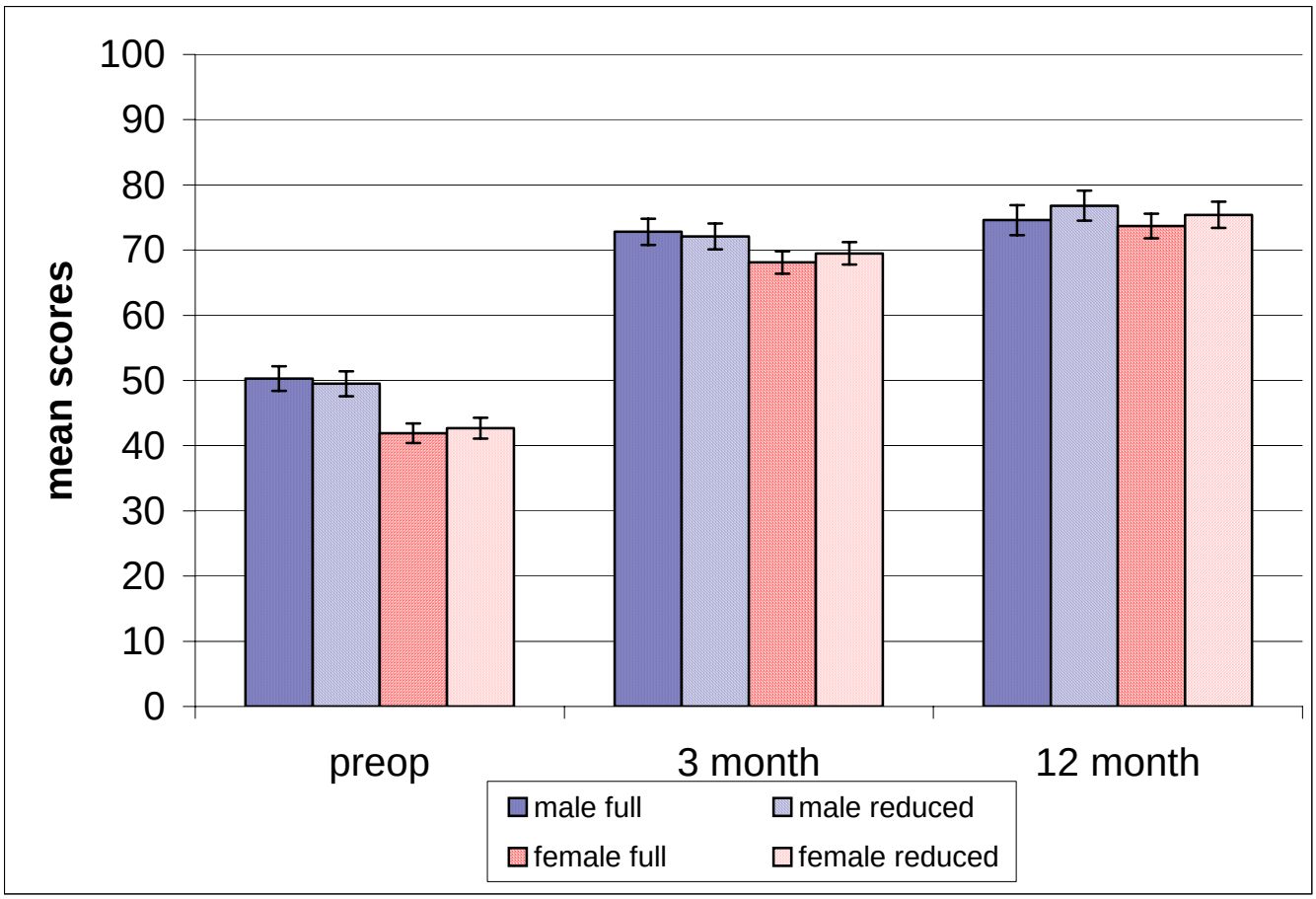

Figure I. Mean WOMAC function scores for the TKR patients stratified by sex with 95% confidence intervals.

References

1. **Bellamy N, Buchanan WW, Goldsmith CH, 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. **Ware JE, Snow K, Kosinski M, Gandek B**. SF-36 health survey manual and interpretation guide. Boston MA. The Health Institute, New England Medical Centre, 1993

4. **Hawker G, Melfi C, Paul J, Green R, Bombardier C**. Comparison of a generic (SF-36) and a disease specific (WOMAC) (Western Ontario and McMaster Universities Osteoarthritis Index) instrument in the measurement of outcomes after knee replacement surgery. *J Rheumatol 1995;22(6):1193-6*

5. **Ware JE, Kosinski M, Keller SD**. A SF-12: an even shorter health survey. *Med Outcomes Trust Bull 1996;4(1):2*

6. **Jenkinson C, Layte R**. Development and testing of the UK SF-12. *J Health Serv Res Policy 1997;2(1):14-8*

7. **Jenkinson C, Layte R, Jenkinson D, Lawrence K, Peterson S, Paice C, Stradling J.** A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies? *J Public Health Med 1997;19(2):179-86*

8. **Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G**. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. *Arthritis Care and Research 1999;12(5):331-5*

9. **Sun Y, Sturmer T, Gunther KP, Brenner H**. Reliability and validity of clinical outcome measurements of osteoarthritis of the hip and knee - a review of the literature*. Clin Rheumatol 1997;16(2):185-98*

10. **Kreibich DN, Vaz M, Bourne RB, Rorabeck CH, Kim P, Hardie R, Kramer J, Kirkley A.** What is the best way of assessing outcome after total knee replacement? *Clin Orthop Rel Res 1996;331:221-5*

11. **Streiner DL, Norman GR**. Health measurement scales. A practical guide to their development and use. *Second edition. Oxford University Press. 2000.*

12. **Liang MH, Fossel AH, Larson MG.** Comparisons of five health status instruments for orthopaedic evaluation. *Med Care 1990;28(7):632-642*

13. **Insall JN, Dorr LD, Scott RD, Scott WN.** Rationale of the Knee Society clinical rating system. *Clin Orthop Rel Res 1989;248:13-4*

14. **Mahomed N Daltroy L, Fossel AH, Katz JN**. Self administered patient satisfaction scale for joint replacement arthroplasty. 52nd Annual Canadian Orthopaedic Association Meeting, Hamilton, Ontario. 1997

15. **Dawson J, Fitzpatrick R, Murray D, Carr A**. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg [Br] 1998;80-B(1):63-9*

16. **Harris WH.** Traumatic arthritis of the hip after dislocation and acetabular fractures: Treatment by Mold arthroplasty. *J Bone Joint Surg [Am] 1969;60-A:454-6*

17. **Kleinbaum DG, Kupper LL, Muller KE**. Testing for the equality of two correlations. *Applied Regression Analysis and Other Multivariate Methods. Belmont, California, Duxbury Press. 1988;91-3*

18. **Fortin PR, Stucki G, Katz JN.** Measuring relevant change: an emerging challenge in rheumatologic clinical trials. *Arthritis Rheum 1995;38(8):1027-30*

19. **Harcourt WGV, White SH, Jones P.** Specificity of the Oxford knee status questionnaire. *J BoneJoint Surg [Br] 2001;83-B:345-7*

20. **Taylor AH, Pattison G, Whitehouse SL, Blom A, Bannister GC.** Postal questionnaires – do they work? *Hip International 2000;10(4):251*