Osteoarthritis and Cartilage Open 4 (2022) 100322

Contents lists available at ScienceDirect

Osteoarthritis and Cartilage Open

journal homepage: www.elsevier.com/journals/osteoarthritis-and-cartilage-open/2665-9131

Validation and interval scale transformation of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) in patients undergoing knee arthroplasty, using the Rasch model

Ying-Ying Leung [a,*], Julian Thumboo [a], Seng-Jin Yeo [b], Vikki Wylde [c], Alan Tannant [d]

[a] Department of Rheumatology & Immunology, Singapore General Hospital, Singapore
[b] Department of Orthopedic Surgery, Singapore General Hospital, Singapore
[c] Musculoskeletal Research Unit, Bristol Medical School, University of Bristol, UK
[d] Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, UK

ARTICLE INFO

Keywords:
Knee osteoarthritis
WOMAC
Validation
Interval scale measurement

ABSTRACT

Objectives: Interval scale reduce measurement bias compared to ordinal scale. We aimed to evaluate the fit of Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) to the Rasch model and derive the transformation table for interval scale measurement.
Methods: Data from osteoarthritis patients listed for knee arthroplasty (KA) pre-operatively, and at 6- and 12-months post-operative was used. WOMAC was calibrated for fit to the Rasch model for monotonicity, homogeneity, local item independence and absence of differential item functioning (DIF) in a randomly selected 900 patients, 300 from each time point; parameter estimates were then imported into the full data set. Responsiveness was reported through Standard Error of Measurement (SEM); Smallest Detectable Difference (SDD), %SDD and effect sizes (ES) between baseline and 6-months. WOMAC was transformed from ordinal to interval values.
Results: 1136 patients (mean age 65.9 years, 69.9% female) were included. WOMAC pain (0–20), function (0–68) and total scores (0–96) had adequate fit to Rasch model with good reliability (Person Separation Index: 0.76, 0.80 and 0.79). No item deletion was required. The SEM, SDD, %SDD and ES of WOMAC total were 4.4, 6.9, 10.1, and 1.97. No significant DIF was seen for age, sex, body mass index, type of KA, languages, and education level. WOMAC pain, function and total scores were transformed to interval scales.
Conclusion: WOMAC total, pain and function scales had adequate fit to the Rasch model, providing unidimensional measure with good reliability and responsiveness. Transformation of WOMAC to interval scale measurement is applicable to other studies.

1. Introduction

Knee osteoarthritis (OA) is prevalent and is a major cause of disability worldwide [1]. An early population study found that prevalence of knee problems was 17% in those age 55–64, rising to 32% in those aged 85 years and above [2], most commonly caused by knee OA. Studies that have defined knee OA by radiography or other classifications have found a similar prevalence depending on the age range and the sex distribution of the population studied [3,4]. Sex-specific risk factors for knee OA have been observed, especially body mass index (BMI), which is reported to affect females more than males [5,6].

The most frequently used Patient-Reported Outcome Measure (PROM) in OA [7] is the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [8]. It is frequently used to determine the improvement following knee arthroplasty (KA), both for OA and rheumatoid arthritis [9,10] or, for example, for various interventions to help improve rehabilitation following KA [11,12]. While early systematic reviews of outcome measures, including the WOMAC, found some with adequate validity and reliability, they reported at that time none of the outcome measures had been comprehensively tested across all clinimetric properties [13]. A more recent systematic review of PROMS used following KA concluded that the instruments assessed, including the WOMAC, had insufficient evidence for their validity in that context [14]. This highlights the need for research to understand the measurement properties of joint-specific PROMs to inform clinical trials and observational studies. Another systematic review raised the question of the lack

* Corresponding author. Department of Rheumatology and Immunology, Singapore General Hospital, The Academia, level 4, 20 College Road 169856, Singapore.
E-mail address: katyccc@hotmail.com (Y.-Y. Leung).

https://doi.org/10.1016/j.ocarto.2022.100322
Received 27 April 2022; Received in revised form 9 November 2022; Accepted 10 November 2022

of a ‘total score’ for the WOMAC, with its psychometric evidence focused upon individual subscales [15].

The data collected from WOMAC is usually based on a Likert scale and the scoring of WOMAC subscales and total scores are presented as ordinal scales. Any point change along an ordinal scale is not equal, for example a reduction from 5 points to 4 points, is not equivalent to a reduction from 4 points to 3 points. This violates meaningful measurement that is based on the arithmetical property of interval scales [16,17]. Lack of attention to this important requirement has been shown, for example, to lead to significant under-reporting of patients achieving a Minimally Important Difference (MID) [18]. Fitting the PROMs into Rasch model provides the only mean to convert an ordinal scale to an interval scale. An interval scale or measurement using a linear scale provides the basic requirement for linear measurement over time. Psychometric evidence for the WOMAC has been provided over two decades though classical factor analytic techniques [19] and by fit of its data to the Rasch Measurement model [20–23]. However, previous results applying the Rasch model were not necessarily consistent. For example, some items were removed due to misfit to the model [23], or lacking item stability over time [24]. The application of Rasch analysis has since developed with some major changes to the way in which data are assessed by the model [25]. At the meeting of the Outcome Measure of Rheumatology (OMERACT 11), a special interest group (SIG) on Rasch model analysis emphasized the importance of interval measurement [26]. Yet, the lack of availability of transformation tables limited the application of it in practice. Therefore, this study aimed to review fit of the WOMAC data to the Rasch model in a large sample of patients with knee OA who have undergone KA, and to develop the necessary transformation table for interval measurement.

2. Methods

2.1. Patients and setting

Patients listed for KA due to knee OA were recruited from May 2017 to May 2019. Patients were recruited 2 weeks prior to their scheduled surgery, in the pre-operative assessment clinic of a single tertiary referral centre, Singapore General Hospital, Singapore. The study protocol was approved by the Singhealth Central Institution Review Board (CIRB Ref 2016/3168E) and all patients provided informed, written consent prior to participation.

2.2. Data collection

Data were collected pre-operatively, and at 6- & 12-months post-KA. Data collected included demographics and educational status, work status, type of KA (total KA versus uni-compartmental KA), and condition of contralateral knee in categories (e.g. unilateral KA, other knee asymptomatic). Height and weight were measured to calculate body mass index (BMI). Comorbidities was self-reported using the Functional Comorbidity Index for 16 conditions, over and above their arthritis [27]. Self-reported PROMs were completed in paper and pencil format in either Chinese or English, depending on patients' preference.

The key PROM in test, WOMAC, consists of three domains [8], pain [5 items], stiffness [2 items] and physical function [17 items]. Items are presented as a Likert style response option with five categories, ranging from none to extreme. A higher score indicates higher level of pain, stiffness or disability. A total score is suggested by summation of the scores from three domains [28,29], and has been widely used in clinical trials of hip and knee OA [30]. Despite its use, there appears to be little psychometric evidence to-date to support a total score. While the WOMAC has different versions developed over time, and some with different response options, the version used in the current study is 3.0 with the Likert response options. Both the English and Chinese versions of WOMAC were validated for traditional clinimetric properties in knee OA in Singapore [31,32]. WOMAC pain was collected for the index knee, where the index knee was designated as the side undergoing KA, or the dominant side if bilateral KA.

2.3. Statistical analysis

2.3.1. Psychometric evaluation -the Rasch model

Data from the WOMAC were fit to the Rasch model. The three domains of the WOMAC were fit separately, and then as a total set to determine if a total score was viable. Recent applications of the Rasch model emphasized the requirements of i) unidimensionality; ii) monotonicity; iii) homogeneity; iv) local item independence and v) group invariance or absence of differential item functioning (DIF) [33,34]. Set of items to be added together to provide a score had to satisfy all of the above-mentioned requirements [35–38]. The local item independence requirement was tested through an analysis of the residual correlations whereby a correlation value across pairs of items of 0.2 above the average correlation is considered a breach of that requirement [39]. Where dependency was found, testlets (*a priori* grouping) or ‘super items’ (post-hoc grouping) were applied to absorb the dependency [40]. Unidimensionality was tested through the approach indicated by Smith, whereby the principal component analysis of the residuals identified two sets of items with opposite loadings, whose person estimates were then compared, and if less than 5% of these are significant, the scale was deemed unidimensional [41]. Most recently, a bi-factor equivalent solution within the framework of the unidimensional Rasch model has become available which can enable the derivation of a total score in the presence of some multidimensionality. Essentially the approach provides a bi-factor-equivalent solution basing the person estimate on the common variance amongst the subscales [42,43].

Homogeneity was tested through a Chi-Square fit statistic, which examines item hierarchy along groups with different scores, and group invariance or absence of differential item functioning (DIF) by an ANOVA of the residuals across key contextual groups. These groups included time of assessment, age, sex, BMI, type of KA, language of the questionnaire (Chinese and English) and educational level. Invariance (free of DIF) is important to ensure the PROM and subscales are not biased between subgroups of patients for the intended use. The basic DIF evaluation should include age and sex but, in addition, we evaluated the invariance across languages, BMI groups, and type of KA.

Where possible, parallel forms were considered to give access to a conditional chi-square fit statistic in the RUMM2030 software, which is known to provide robust fit statistics [44,45]. Forms were deemed parallel if their latent correlation was 0.9 or above. This was consistent with split half reliability, where values of 0.9 and above were considered essential for comparisons at the individual level [46]. The proportion of variance discarded to make a unidimensional latent estimate under a parallel form condition was also reported.

The analysis was undertaken on a calibration sample of 900 cases where 300 cases were randomly selected from each time point and, to avoid dependency over time, no individual appears in the total sample more than once [47]. Should data fit the model satisfactorily, the parameter estimates obtained from this calibration sample were then imported into the full data set to provide the necessary estimates of all the WOMAC scales. Transformation tables of the ordinal to interval values were constructed.

2.3.2. Scale reliability and responsiveness

Reliability was reported, both as a Peron Separation Index (PSI) in the Rasch software and as internal consistency reliability by Cronbach's alpha. The minimum values of 0.7 and 0.85 indicate acceptable reliability for group and individual use, respectively [48]. Responsiveness was reported through a series of metric-based indicators [49], including the Standard Error of Measurement (SEM); Smallest Detectable Difference (SDD) [50]; %SDD of the full operational range of the scale, and the Effect Size of the differences across the repeated metric measurement of individual scales, in particular for difference between baseline and 6-months.

3. Results

3.1. Demographics and clinical characteristics

One thousand two hundred and one patients with knee OA were recruited and data for 1136 who underwent KA were analysed. Follow-up data was available for 97.1% and 95.8% of patients at 6 and 12 months post-operatively. Their mean age was 65.9 years (standard deviation, SD 7.0), and 794 (69.9%) were female. Just under one-tenth (9.2%) reported living alone, 17.6% had post-secondary education, and 43.3% were working. Over half 56% reported comorbid conditions, along with their arthritis, the most common being visual impairment (cataract, glaucoma, macular disease) (18.8%), diabetes (18.4%) and osteoporosis (13.5%). The mean BMI was 28.3 kg/m^2 and 32.3% had a BMI ≥30 kg/m^2. An overview of participant demographics is provided in Table 1.

Just over one-fifth of patients (20.5%) had a unilateral KA where the other knee was not symptomatic. Almost half (46.9%) had a unilateral KA where the other knee was symptomatic. Just over a quarter (27.2%) had a unilateral KA where the other knee was already replaced. Finally, 5.4% had a bilateral KA. Over three in five (62%) reported that they had pain in the index knee for three or more years.

Table 1
Baseline characteristics of patients listed for and undergone KA.

Baseline characteristics	
Total Sample, n	1136
Age, years[a]	65.9 (7.0)
Female/male sex, n (%)	794 (69.9)/342 (30.1)
Ethnicity, n (%)	
Chinese	954 (84.0)
Malay	84 (7.4)
India	82 (7.2)
Others	16 (1.4)
Highest education n (%)	
None to primary	464 (40.8)
Secondary	472 (41.5)
Post-secondary	129 (11.4)
Tertiary and above	71 (6.3)
Type of KA, n (%)	
Total KA	1045 (92.0)
Uni-compartmental KA	91 (8.0)
Condition of contralateral knee, n (%)	
Unilateral KA, other knee not symptomatic	233 (20.5)
Unilateral KA, other knee symptomatic	533 (46.9)
Unilateral KA, other knee replaced	309 (27.2)
Bilateral KA	61 (5.4)
Duration of knee pain on index knee, n (%)	
Less than one year	150 (13.2)
1–2 years	227 (20.0)
3–5 years	332 (29.2)
6–8 years	150 (13.2)
9–12 years	135 (11.9)
13–15 years	36 (3.2)
Over 15 years	106 (9.3)
Living alone, n (%)	105 (9.2)
Working status, n (%)	
Working full time/part time	493 (43.3)
Not working by choice	552 (48.6)
Unable to work due to knee OA	89 (7.8)
Unable to work due to other health conditions	2 (0.2)
Looking for job	1 (0.1)
BMI, kg/m^2[a]	28.3 (4.7)
BMI ≥30 kg/m^2, n (%)	(32.3)
Top 5 self-reported comorbidities beyond arthritis, n (%)	
Visual impairment (cataract, glaucoma, macular disease)	213 (18.8)
Diabetes mellitus	209 (18.4)
Osteoporosis	153 (13.5)
Degenerative disc disease (back, spinal stenosis, severe chronic back pain)	134 (11.8)
Upper gastrointestinal disease (ulcer, hernia, reflux)	95 (8.4)

BMI: body mass index; KA: knee arthroplasty; n: number; OA: osteoarthritis; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.
[a] Mean (SD) unless specified.

Table 2
WOMAC scores for patients undergone KA.

	Baseline	6-month	12-month
Sample, n	1136	1103	1088
Raw scores			
WOMAC pain, 0-20	6.8 (3.3)	1.4 (2.2) **	0.9 (2.1) **
WOMAC stiffness, 0-8	2.3 (1.8)	1.1 (1.2) **	0.7 (1.1) **
WOMAC function, 0-68	19.4 (10.6)	5.1 (6.5) **	3.6 (6.6) **
WOMAC total, 0-96	28.4 (14.3)	7.6 (9.0) **	5.2 (9.0) **
Interval Scales			
WOMAC pain, 0-20	8.0 (2.9)	2.1 (2.7) **	1.3 (2.5) **
WOMAC function, 0-68	35.8 (10.1)	15.3 (11.4) **	10.5 (11.8) **
WOMAC total, 0-96	30.3 (6.5)	14.9 (9.1) **	10.6 (9.8) **

All data given as mean (SD). Wilconxon signed rank tests and *t*-test were made in comparison to baseline scores for raw score and interval scales respectively.
*$p < 0.05$; **$P < 0.01$.
KA: knee arthroplasty; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

There was a significant difference in age according to the type of KA, ranging from 63.6 years for those with a bilateral KA to 67.5 years for those who had a unilateral KA where the other knee was already replaced (F11.35 (df3); p=<0.001). There was no difference by sex (Chi-Square 5.202 (df3); p = 0.158). There was however a significant difference across the type of KA for those with a BMI ≥30 kg/m^2. This ranged from 23.6% of those who had a unilateral KA where the other knee was not symptomatic, to 47.5% of those who had a bilateral KA (Chi-Square 18.6 (df3); p= <0.001). The study cohort had moderate knee OA symptoms. The mean (SD) WOMAC pain, function and total were 6.8 (3.3), 19.4 (10.6) and 28.4 (14.3) respectively. The WOMAC scores of the cohort at baseline, 6-month and 12-month are shown in Table 2. Statistically significant improvement in WOMAC pain, function and total scores were seen at 6-month and 12-month post-operatively.

3.2. Rasch analysis

Initially a likelihood ratio test was undertaken to ascertain which form of the polytomous Rasch model should be used. A significant difference indicated that the partial credit parameterisation of the model should be used. Fit of the Pain subscale in the calibration sample is shown in Table 3-Analysis 1. Fit was adequate, with good reliability and unidimensionality. The transition (threshold) from 'none' to 'mild' for the item 'pain going up or down stairs' was the easiest to affirm. The transition from 'severe' to 'extreme' for the item 'pain sitting or lying' was the most difficult to affirm. The items 'pain at night while in bed' and 'pain sitting or lying' were locally dependent, and thus merged into a super item. The remaining three items were also merged to gain access to the conditional Chi-Square test of fit. The scale was invariant to all contextual factors except time, where pre-operative values were slightly (but significantly -t-test <0.05) higher than post-surgery at any given level of pain. After splitting a super item for time, the effect size of the difference in person estimates between unsplit and split solutions showed a trivial effect size of 0.098, and so no further action was taken. No other significant DIF was seen with age, sex, language, BMI, type of KA and educational level.

The overall distribution showed a significant skew to no pain, but this masked the fact that this was dominated by the 12-month assessment, and to a lesser extent by six months, whereas pre-operative dominated higher levels of pain. The effect sizes for differences in level of pain in this calibration sample, from pre-operative to six months, pre-operative to 12 months, and six months to 12 months were 2.08, 2.56 and 0.371 respectively, suggesting significant reduction in pain for the first six months, with some continuing improvement until 12 months.

Table 3
Fit of WOMAC scales to the Rasch model.

Analysis	Scale/Construct	Residuals		Conditional Chi-Square		Reliability		Unidimens-ionality	ECV	Latent Corr
	WOMAC	Item	Person	Value (df)	P	PSI	α	% Significant t-tests		
Calibration Sample (n = 900)										
1	Pain	2.499	0.779	16.6 (11)	0.121	0.75	0.82	1.57	0.93	0.93
2	Stiffness	1.215	0.462	57.0 (9)[b]	<0.001	0.60	0.82	2.53^	–	–
3	Function	2.385	0.849	53.3 (42)	0.113	0.87	0.94	1.16	0.96	0.93
4	Total	0.511	0.858	71.6 (60)	0.145	0.90	0.80	2.30	0.98	095
Baseline, total sample (n = 1136)										
5	Pain	5.134	0.8826	16.8 (11)	0.114	0.80	0.74	4.8	0.97	0.99
6	Stiffness	1.183	0.621	74.1 (10)[b]	<0.001	0.69	0.81	1.3^	–	–
7	Function	1.076	1.006	34.2 (48)	0.933	0.91	0.92	4.3	1.00	1.00
8	Total	0.063	0.960	76.9 (67)	0.191	0.91	0.92	4.0	0.98	0.97
6 Months, total sample (n = 1102)										
9	Pain	1.721	0.826	13.3 (8)	0.102	0.40	0.83	1.62	0.91	0.88
10	Stiffness	0.790	0.461	127.3 (6)[b]	<0.001	0.48	0.74	1.24^	–	–
11	Function	2720	0.774	34.8 (34)	0.4228	0.81	0.92	3.7	0.99	1.00
12	Total	4.566	0.639	53.7 (39)	0.058	0.65	0.75	2.4	0.79	0.93
12 months, total sample (n = 1089)										
13	Pain	0.638	0.621	8.4 (8)	0.395	0.291	0.89	1.76	NA	NA
14	Stiffness	0.330	0.535	95.8 (6)[b]	<0.001	0.16	0.72	2.2^	–	–
15	Function	3.537	0.720	28.9 (31)	0.574	0.69	0.92	1.5	1.00	0.99
16	Total	2.598	0.666	53.0 (42)	0.119	0.69	0.94	17	0.98	0.97
	Ideal value	<1.4[a]	<1.4		>0.01[¥]			<5.0	≥0.9	>0.9

¥ Bonferroni adjusted value; ^ Under-powered; PSI: Person Separation Index; α: Cronbach's Alpha; ECV: Explained Common Variance; Latent Corr: the correlation between parallel forms; NA Not Available; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.
[a] Only with equal size super items/testlets, else inflated.
[b] Ordinary Chi-Square.

Fit of the Stiffness subscale was poor, and with just two items, little could be done to redress this situation (Table 3-Analysis 2). The item 'How much stiffness do you have after sitting, lying or resting later in the day' showed DIF by time. Splitting this item by time did not resolve the misfit.

In the Function subscale, the transition from 'none' to 'mild' for the item 'difficulty with descending stairs' was the easiest to affirm. The transition from 'severe' to 'extreme' for the item 'difficulty rising from bed' was the most difficult to affirm. Clusters of locally dependent items were found throughout this scale. The items 'difficulty putting on socks/stockings' and 'difficulty taking off socks/stockings; had a residual correlation of 0.937, suggesting that one of these items was completely redundant. The items 'difficulty with descending stairs' and 'difficulty with ascending stairs' had a residual correlation of 0.523. These patterns of dependency were used to make two parallel forms with a latent correlation of 0.92, which showed adequate fit to the model (Table 3-Analysis 3). The scale was invariant (free of DIF) to all contextual factors except time and language. For time, 6-month post-operative values were significantly different (-t-test <0.05) to both pre-and 12 months values at any given level of functioning. After splitting a parallel form for time, the effect size of the difference in person estimates between unsplit and split solutions showed a small effect size of 0.402. After splitting the parallel forms for language, giving a significant difference between the split and unsplit solutions, the effect size was found to be 0.01, considered trivial, and no action was taken.

Finally, a total score was considered. A conceptual basis was chosen for this, deriving two testlets, one for the pain and stiffness subscales, the other from the function subscale. Fit to the model was adequate (Table 3-Analysis 5). DIF was only observed for time on the pain and stiffness testlet. This testlet was split with times at pre-operative and 6-month versus time at 12-month post-operatively. While the *t*-test between the spit and unsplit solutions was significant (p= <0.001), the Effect Size of the difference in estimates was just 0.026, and so no further action was taken. This bi-factor equivalent solution retained 90% of the variance. Accommodating the local dependency amongst the item set gave a (α) reliability of 0.80.

Results of fitting the data to the model for each domain at each individual time point is also shown in Table 3 (Analyses 5–16). The mean (SD, interquartile range) WOMAC total score (interval scale) at baseline, 6 and 12-months were 30.3 (6.5, 27.0–34.2), 14.9 (9.1, 9.3–20.8) and 10.6 (9.8, 0.0–16.4), respectively (Table 2).

3.3. *Scale reliability and responsiveness*

The person separation index (PSI) of the WOMAC pain, function and total scores were 0.76, 0.80 and 0.79 respectively (Table 3). The WOMAC Pain subscale showed the largest effect size for the change between pre-operative and 6 months (Table 4). The WOMAC Function and Total score had the smallest percent of their total operational range to negotiate to get above the error (%SDD), although this was similar for all the (sub)

Table 4
Scaling and responsiveness of the WOMAC, Oxford Knee Score and the SF-36 Physical Functioning (PF) score.

(Sub)Scale	Ordinal Scaling (Baseline n = 1136)			Interval Scaling (Baseline n = 1136)						Effect Size
	Reliability (Cronbach's α)	Median	IQR	Reliability (PSI)	Mean	SD	SEM	SDD	%SDD	T1-T2
WOMAC										
Pain	0.87	7	4–9	0.76	8.0	2.9	1.04	2.90	14.5	2.11
Stiffness	082	2	1–4	0.60	–	–	–	–	–	–
Physical Function	0.94	18	12–26	0.80	35.7	10.2	2.49	6.90	10.1	1.91
Total	0.80	27	18–37	0.79	44.2	9.9	4.4	6.9	10.1	1.97

*Ordinal items scored as 1–5, range 12–60 Metric scored 0–4, range 0–48; **Ordinal Items scored as 1–3, range 10–30, metric 0–2 range 0–20.
IQR: Inter-Quartile Range; PSI: Person Separation Index; SD: Standard Deviation; SEM: Standard Error of measurement; SDD: Smallest Detectable Difference; T1: Pre-operative, T2: 6-Months; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

Table 5
Transformation of Ordinal raw score to Interval measure for Western Ontario and McMaster Universities Osteoarthritis Index.

Raw Score	Interval scale		
	Total	Pain	Function
0	0.0	0.0	0.0
1	5.6	2.1	6.5
2	9.3	3.6	11.0
3	11.7	4.8	14.1
4	13.6	5.7	16.6
5	15.1	6.7	18.7
6	16.4	7.5	20.7
7	17.6	8.3	22.5
8	18.7	9.1	24.2
9	19.8	9.8	25.8
10	20.8	10.6	27.3
11	21.7	11.3	28.7
12	22.6	12.1	30.1
13	23.4	12.9	31.5
14	24.2	13.6	32.7
15	24.9	14.4	34.0
16	25.7	15.1	35.1
17	26.3	15.8	36.2
18	27.0	16.8	37.3
19	27.6	18.1	38.3
20	28.2	20.0	39.2
21	28.7		40.1
22	29.2		41.0
23	29.7		41.8
24	30.2		42.6
25	30.6		43.3
26	31.0		44.0
27	31.4		44.6
28	31.8		45.1
29	32.1		45.7
30	32.4		46.2
31	32.7		46.7
32	33.0		47.1
33	33.3		47.5
34	33.5		47.8
35	33.8		48.2
36	34.0		48.5
37	34.2		48.8
38	34.4		49.1
39	34.7		49.3
40	34.9		49.6
41	35.1		49.8
42	35.3		50.0
43	35.5		50.2
44	35.7		50.4
45	35.9		50.6
46	36.1		50.7
47	36.3		50.9
48	36.6		51.1
49	36.8		51.3
50	37.0		51.5
51	37.3		51.6
52	37.6		51.8
53	37.9		52.0
54	38.2		52.2
55	38.5		52.4
56	38.8		52.6
57	39.2		52.9
58	39.5		53.1
59	39.9		53.5
60	40.4		53.8
61	40.8		54.3
62	41.3		54.9
63	41.8		55.6
64	42.3		56.6
65	42.8		58.0
66	43.4		60.0
67	44.0		63.1
68	44.6		68.0
69	45.3		
70	46.0		
71	46.7		
72	47.5		
73	48.3		
74	49.2		
75	50.0		
76	51.0		
77	52.0		
78	53.0		
79	54.1		
80	55.2		
81	56.5		
82	57.8		
83	59.2		
84	60.7		
85	62.3		
86	64.1		
87	66.1		
88	68.3		
89	70.6		
90	73.2		
91	75.9		
92	78.9		
93	82.0		
94	85.5		
95	90.0		
96	96.0		

scales, but the WOMAC Pain scale was the most challenged in this respect.

3.4. Transformation table

A transformation table for the WOMAC ordinal to interval scores of the Total, Pain and Function scales is provided in Table 5. The distribution of changes of WOMAC total scores in interval scale at 6-month from baseline are shown in Fig. 1, illustrating the normally distribution of the change scores.

4. Discussion

In this study, we demonstrated the construct validity, internal consistency reliability and responsiveness of WOMAC using Rasch analysis in

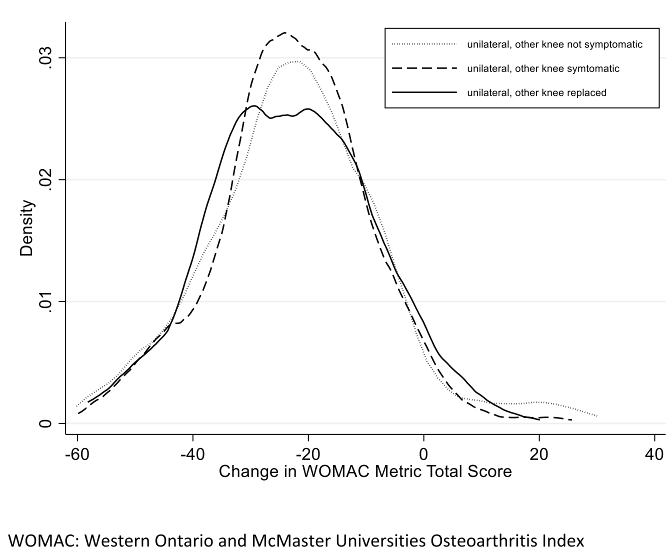

Fig. 1. Kernal Density Graph of WOMAC Total Score distributions of change from 0 to 6 months by KA Category. WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

a large sample of patients with knee OA listed for KA. The key deliverable of the current study is that we derived the transformation table, allowing transformation of the WOMAC pain, function, and total from ordinal into interval scales. This transformation is applicable for WOMAC scores collected in all studies involving patients with OA, given the large sample size we used for cross sectional and longitudinal validation. Clinicians and researchers can now convert the WOMAC ordinal scale into interval scales at ease using the transformation table. The transformation of ordinal to equal interval scaling is fundamental to meaningful measurement based on arithmetical property and is essential to accurately informing the magnitude of change and the reporting of responsiveness [16,17].

The WOMAC Pain and Function subscales were shown to fit the Rasch model, and the two stiffness items were incorporated into the total score, achieved through a bi-factor equivalent solution, retaining 90% of the variance that was common. The level of reliability of the WOMAC subscales in the current study were smaller than that previously observed [51]. This may well be due to the accommodation of local item dependency in the current analysis, which is known to inflate reliability [52,53]. This can be observed by the comparison between the ordinal Cronbach's alpha, where no adjustment is made, and the interval-bases Person Separation Index. No items were deleted under this solution, which has the advantages of maintaining comparability with much of the published literature. However, this is only under the condition of the transformation of the ordinal total score to the interval scaling, and the required discarding of unique variance. Responsiveness of WOMAC pain, function and total scores were excellent as shown by the effect sizes for the difference between pre-operative and the 6-month assessment.

The ability to obtain a total score for the WOMAC with a bi-factor equivalent solution reflects some earlier evidence that, at least, the Pain and Function subscales may represent the same construct [21,51]. This may be related to the fact that the pain questions are not primarily about the degree of pain, but rather contextualising the pain experience when undertaking tasks (e.g. pain interference), whereas the Function scale is about the degree of difficulty in performing tasks. Tasks are quite similar, for example ‘going up and down stairs’ in the Pain subscale, separated into ascending and descending stairs in the Function subscale. The WOMAC total score was developed from its original version to provide an easy interpretation for trials [8]. Although reporting the pain and function sub-domains were more commonly used, WOMAC total has been used in some studies and the need for a total score has been highlighted [11]. Our current data should provide psychometric support and transformation table for its use as an interval scale. Future work should explore the influence of expectations and mental health, upon both indication for, and result of KA [54].

It worth to mention about the limitations for this study. We derived the WOMAC interval scales from a cohort of patients with advanced stage knee OA listed for KA. The results may not be generalized to patients with earlier stages of knee OA. However, we have collected data both pre-operatively as well as 6 and 12-months post-operatively when most patients had a moderate to significant great improvement, which does cover a wide of severity in patients with knee OA. Secondly, we derived the WOMAC interval scales using WOMAC 3.0 version with the Likert response options, which may not apply to other versions. The total score can only be used with complete data, although it has been shown that single imputation has little effect on interpreting fit to the Rasch model (47). Lastly, we did not collect data repeatedly before a change should happen and thus cannot report test-retest reliability in this paper, although this has been well-reported previously [55,56].

The strengths of the study include the large sample size, including a good distribution of the type of KR; the application of the Rasch model to deliver valid interval levels estimates for group comparisons and calculating change scores, and the confirmation that a total WOMAC score is viable under a bi-factor solution. With these, we provided the transformation table, allowing easy conversion of WOMAC pain, function, and total scores from ordinal into interval scales.

In conclusion, the WOMAC total, pain and function subscales had adequate fit to the Rasch model, providing unidimensional measure with good reliability and responsiveness. Transformation of WOMAC to interval scale measurement is feasible and applicable to wide range of studies involving patients with knee OA.

Funding

This study was supported by the National Medical Research Council of Singapore (NMRC/HSRG/0061/2016 and NMRC/CSA-INV/0022/2017).

Author contributions

The authors declare the following contributions to the preparation of manuscript: study conception and design (YYL and AT), acquisition of data (YYL, JT, SJY), data analysis (AT), interpretation of data (all authors), drafting of manuscript (YYL and AT), critical revision of manuscript for important intellectual content (all authors), and final approval of the manuscript (all authors). YYL (katyccc@hotmail.com) takes responsibility for the integrity of the work as a whole.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

NA.

References

[1] T. Vos, A.D. Flaxman, M. Naghavi, R. Lozano, C. Michaud, M. Ezzati, et al., Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010, Lancet 380 (2012) 2163–2196.

[2] E.M. Badley, A. Tennant, Changing profile of joint disorders with age: findings from a postal survey of the population of Calderdale, West Yorkshire, United Kingdom, Ann. Rheum. Dis. 51 (1992) 366–371.

[3] J.W. Hong, J.H. Noh, D.J. Kim, The prevalence of and demographic factors associated with radiographic knee osteoarthritis in Korean adults aged >/= 50 years: the 2010-2013 Korea National Health and Nutrition Examination Survey, PLoS One 15 (2020), e0230613.

[4] Y.Y. Leung, S. Ma, M. Noviani, S.B. Wong, C.M. Lee, I.A. Soh, et al., Validation of screening questionnaires for evaluation of knee osteoarthritis prevalence in the general population of Singapore, Int. J. Rheum. Dis. 21 (2018) 629–638.

[5] I.A. Szilagyi, J.H. Waarsing, D. Schiphof, J.B.J. van Meurs, S.M.A. Bierma-Zeinstra, Towards sex-specific osteoarthritis risk models: evaluation of risk factors for knee osteoarthritis in males and females, Rheumatology 61 (2) (2021) 648–657, https://doi.org/10.1093/rheumatology/keab378.

[6] Y.Y. Leung, J.C. Allen Jr., M. Noviani, L.W. Ang, R. Wang, J.M. Yuan, et al., Association between body mass index and risk of total knee replacement, the Singapore Chinese Health Study, Osteoarthritis Cartilage 23 (2015) 41–47.

[7] Lundgren-Nilsson Å, A. Dencker, A. Palstam, G. Person, M.C. Horton, R. Escorpizo, et al., Patient-reported outcome measures in osteoarthritis: a systematic search and review of their use and psychometric properties, RMD Open 4 (2018), e000715.

[8] N. Bellamy, W.W. Buchanan, C.H. Goldsmith, J. Campbell, L.W. Stitt, 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 (1988) 1833–1840.

[9] I. Qadir, L. Khan, J. Mazari, U. Ahmed, A.U. Zaman, A. Aziz, Comparison of functional outcome of simultaneous and staged bilateral total knee arthroplasty : systematic review of literature, Acta Orthop. Belg. 87 (2021) 487–493.

[10] A.A. Küçükdeveci, A.H. Elhan, B.D. Erdoğan, Ş. Kutlay, D. Gökmen, C. Ateş, et al., Use and detailed metric properties of patient-reported outcome measures for rheumatoid arthritis: a systematic review covering two decades, RMD Open 7 (2021).

[11] D. Wang, T. Wu, Y. Li, L. Jia, J. Ren, L. Yang, A systematic review and meta-analysis of the effect of preoperative exercise intervention on rehabilitation after total knee arthroplasty, Ann. Palliat. Med. 10 (2021) 10986–10996.

[12] W. Alrawashdeh, J. Eschweiler, F. Migliorini, Y. El Mansy, M. Tingart, B. Rath, Effectiveness of total knee arthroplasty rehabilitation programmes: a systematic review and meta-analysis, J. Rehabil. Med. 53 (2021), jrm00200.

[13] T.E. Howe, L.J. Dawson, G. Syme, L. Duncan, J. Reid, Evaluation of outcome measures for use in clinical practice for adults with musculoskeletal conditions of the knee: a systematic review, Man. Ther. 17 (2012) 100–118.
[14] S.A. Sabah, E.A. Hedge, S.G.F. Abram, A. Alvand, A.J. Price, S. Hopewell, Patient-reported outcome measures following revision knee replacement: a review of PROM instrument utilisation and measurement properties using the COSMIN checklist, BMJ Open 11 (2021), e046169.
[15] Y. Wang, M. Yin, S. Zhu, X. Chen, H. Zhou, W. Qian, Patient-reported outcome measures used in patients undergoing total knee arthroplasty, Bone Joint Res 10 (2021) 203–217.
[16] B.D. Wright, J.M. Linacre, Observations are always ordinal; measurements, however, must be interval, Arch. Phys. Med. Rehabil. 70 (1989) 857–860.
[17] C. Merbitz, J. Morris, J.C. Grip, Ordinal scales and foundations of misinference, Arch. Phys. Med. Rehabil. 70 (1989) 308–312.
[18] L.M. Ornbjerg, K.B. Christensen, A. Tennant, M.L. Hetland, Validation and assessment of minimally clinically important difference of the unadjusted Health Assessment Questionnaire in a Danish cohort: uncovering ordinal bias, Scand. J. Rheumatol. 49 (2020) 1–7.
[19] F. Angst, T. Ewert, S. Lehmann, A. Aeschlimann, G. Stucki, The factor subdimensions of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) help to specify hip and knee osteoarthritis. a prospective evaluation and validation study, J. Rheumatol. 32 (2005) 1324–1330.
[20] G. Rasch, Probabilistic models for some intelligence and attainment tests, in: (Copenhagen, Danish Institute for Educational Research), The University of Chicago Press, Chicago, 1960.
[21] L. Ryser, B.D. Wright, A. Aeschlimann, S. Mariacher-Gehler, G. Stucki, A new look at the western Ontario and McMaster Universities osteoarthritis index using Rasch analysis, Arthritis Care Res. 12 (1999) 331–335.
[22] F. Wolfe, S.X. Kong, Rasch analysis of the Western Ontario MacMaster questionnaire (WOMAC) in 2205 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia, Ann. Rheum. Dis. 58 (1999) 563–568.
[23] A.M. Davis, E.M. Badley, D.E. Beaton, J. Kopec, J.G. Wright, N.L. Young, et al., Rasch analysis of the Western Ontario McMaster (WOMAC) Osteoarthritis Index: results from community and arthroplasty samples, J. Clin. Epidemiol. 56 (2003) 1076–1083.
[24] Y. Ko, N.N. Lo, S.J. Yeo, K.Y. Yang, W. Yeo, H.C. Chong, et al., Rasch analysis of the Oxford knee score, Osteoarthritis Cartilage 17 (2009) 1163–1169.
[25] Lundgren Nilsson Å, A. Tennant, Past and present issues in Rasch analysis: the functional independence measure (FIM™) revisited, J. Rehabil. Med. 43 (2011) 884–891.
[26] Y.Y. Leung, M.E. Png, P. Conaghan, A. Tennant, A systematic literature review on the application of Rasch analysis in musculoskeletal disease – a special interest group report of OMERACT 11, J. Rheumatol. 41 (2014) 159–164.
[27] D.L. Groll, T. To, C. Bombardier, J.G. Wright, The development of a comorbidity index with physical function as the outcome, J. Clin. Epidemiol. 58 (2005) 595–602.
[28] N.F. Woolacott, M.S. Corbett, S.J. Rice, The use and reporting of WOMAC in the assessment of the benefit of physical therapies for the pain of osteoarthritis of the knee: findings from a systematic review of clinical trials, Rheumatology 51 (2012) 1440–1446.
[29] L. Pulik, K. Romaniuk, K. Jaskiewicz, P. Wojtynski, P. Legosz, P. Maldyk, An update on joint-specific outcome measures in total hip replacement, Reumatologia 58 (2020) 107–115.
[30] B. Copsey, J.Y. Thompson, K. Vadher, U. Ali, S.J. Dutton, R. Fitzpatrick, et al., Problems persist in reporting of methods and results for the WOMAC measure in hip and knee osteoarthritis trials, Qual. Life Res. 28 (2019) 335–343.
[31] J. Thumboo, L.H. Chew, C.H. Soh, Validation of the western Ontario and Mcmaster university osteoarthritis index in asians with osteoarthritis in Singapore, Osteoarthritis Cartilage 9 (2001) 440–446.
[32] F. Xie, S.C. Li, R. Goeree, J.E. Tarride, D. O'Reilly, N.N. Lo, et al., Validation of Chinese western Ontario and McMaster Universities osteoarthritis index (WOMAC) in patients scheduled for total knee replacement, Qual. Life Res. 17 (2008) 595–601.
[33] J.-E. Gustafsson, Testing and obtaining fit of data to the Rasch model, Br. J. Math. Stat. Psychol. 33 (1980) 205–233.
[34] J.A. Teresi, M. Kleinman, K. Ocepek-Welikson, Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures, Stat. Med. 19 (2000) 1651–1683.
[35] L.L. Thurstone, Attitudes can Be measured, Am. J. Sociol. 33 (1928) 529–554.
[36] H.-A. Kang, Y.-H. Su, H.-H. Chang, A note on monotonicity of item response functions for ordered polytomous item response theory models, Br. J. Math. Stat. Psychol. 71 (2018) 523–535.
[37] J. Rost, An unconditional likelihood ratio for testing item homogeneity in the Rasch model, Educ. Res. Perspect. 9 (1982) 7–17.
[38] M. Wilson, Detecting and interpreting local item dependence using a fannily of Rasch models, Appl. Psychol. Meas. 12 (1988) 353–364.
[39] K.B. Christensen, G. Makransky, M. Horton, Critical values for yen's Q3: identification of local dependence in the Rasch model using residual correlations, Appl. Psychol. Meas. 41 (2017) 178–194.
[40] H. Wainer, G.L. Kiely, Item clusters and computerized adaptive testing: a case for testlets, J. Educ. Meas. 24 (1987) 185–201.
[41] E.V. Smith, Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals, J. Appl. Meas. 3 2 (2002) 205–231.
[42] A. Rodriguez, S.P. Reise, M.G. Haviland, Evaluating bifactor models: calculating and interpreting statistical indices, Psychol. Methods 21 (2016) 137–150.
[43] D. Andrich, Components of variance of scales with a bifactor subscale structure from two calculations of#±, Educ. Meas. 35 (2016) 25–30.
[44] K.B. Christensen, S. Kreiner, Item fit statistics, in: K.B. Christensen, S. Kreiner, M. Mesbah (Eds.), Rasch Models in Heath, London: ISTE, London, 2013, p. 101.
[45] D. Andrich, B. Sheridan, G. Luo, RUMM2020: Rasch Unidimensional Measurement Models [Computer Software], 2003.
[46] J.M. Bland, D.G. Altman, Statistics notes: cronbach's alpha, BMJ 314 (1997) 572.
[47] T. Mallinson, Rasch analysis of repeated measures, Rasch Meas. Trans. 25 (2011).
[48] A. Tennant, P.G. Conaghan, The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum. 57 (2007) 1358–1362.
[49] M.E. Cudkowicz, D. Schoenfeld, L. Williams, Improving the responsiveness of rating scales: the challenge of stepping twice into the same river, Neurology 62 (2004) 1666–1667.
[50] T.J.B. Kropmans, P.U. Dijkstra, B. Stegenga, R. Stewart, L.G.M. de Bont, Smallest detectable difference in outcome variables related to painful restriction of the temporomandibular joint, J. Dent. Res. 78 (1999) 784–789.
[51] B. Gandek, Measurement properties of the western Ontario and McMaster Universities osteoarthritis index: a systematic review, Arthritis Care Res. 67 (2015) 216–229.
[52] W.M. Yen, Scaling performance assessments: strategies for managing local item dependence, J. Educ. Meas. 30 (1993) 187–213.
[53] J.F. Lucke, Rassling the hog": the influence of correlated item error on internal consistency, classical reliability, and congeneric reliability, Appl. Psychol. Meas. 29 (2005) 106–125.
[54] A. Escobar, L. García Pérez, C. Herrera-Espiñeira, F. Aizpuru, C. Sarasqueta, M. Gonzalez Sáenz de Tejada, et al., Total knee replacement: are there any baseline factors that have influence in patient reported outcomes? J. Eval. Clin. Pract. 23 (2017) 1232–1239.
[55] E.M. Roos, S. Toksvig-Larsen, Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement, Health Qual. Life Outcome 1 (2003) 17.
[56] P.T.S. Lage, L.A.C. Machado, S.M. Barreto, R.C. de Figueiredo, R.W. Telles, Measurement properties of Portuguese-Brazil Western Ontario and McMaster Universities osteoarthritis index (WOMAC) for the assessment of knee complaints in Brazilian adults: ELSA-Brasil Musculoskeletal cohort, Rheumatol. Int. 40 (2020) 233–242.