Osteoarthritis and Cartilage

OARSI Osteoarthritis RESEARCH SOCIETY INTERNATIONAL

Rasch analysis of the intermittent and constant osteoarthritis pain (ICOAP) scale

B.J. Moreton †‡*, M. Wheeler †§, D.A. Walsh †§, N.B. Lincoln †‡

† Arthritis Research UK Pain Centre, University of Nottingham, UK
‡ Institute of Work, Health and Organisations, University of Nottingham, UK
§ Academic Rheumatology, City Hospital, University of Nottingham, UK

ARTICLE INFO

SUMMARY

Objectives: The Intermittent and Constant Osteoarthritis Pain (ICOAP) questionnaire was developed to assess two forms of pain reported by people with osteoarthritis: intermittent and constant pain. Studies examining its measurement qualities have provided some support for its use as separate and total scales. However, it has not been previously evaluated using Rasch analysis. The current study examined the fit between data obtained from the ICOAP questionnaire and the Rasch model to determine whether it meets the requirements of interval-level measurement.
Design: ICOAP responses from 175 participants with knee osteoarthritis were collected in a cross-sectional questionnaire study. Participants were recruited from hospital clinics and a group who had taken part in previous research. The questionnaires were completed at home and returned by pre-paid envelope and the data were analysed using RUMM2020.
Results: Fit to the Rasch model was achieved for both the Constant and Intermittent subscales following removal of a small number of items. The Total scale initially resulted in substantial misfit to the model, but fit was improved by removing four items that misfit the model. However, several participants presented with high fit residuals, which is consistent with misfit.
Conclusions: The results support the use of Constant and Intermittent subscales as unidimensional measures of pain. The Total scale can be adapted to improve fit to the Rasch model, but there are concerns over participant misfit.

© 2012 Osteoarthritis Research Society International. Published by Elsevier Ltd.

Introduction

The most important reported symptom of osteoarthritis (OA) is pain[1,2]. Pain is often treated as a single outcome in clinical trials, but the experience of pain in OA is complex and diverse[3]. Pain may vary according to its localisation, periodicity and descriptive quality, reflecting diverse pathophysiological mechanisms[4]. Pain questionnaires used in OA research aim to measure the overall severity of pain and/or the severity or nature of specific dimensions of pain. In general, clinical trials seek to evaluate overall pain experience, whereas mechanistic studies require a more detailed understanding of phenotype. Estimates of overall pain severity may be sought by combining responses to questionnaires that target specific pain dimensions. However, measurement scales should ideally demonstrate unidimensionality, and combining distinct dimensions into a single scale can sometimes threaten validity[5].

The Intermittent and Constant Osteoarthritis Pain (ICOAP) questionnaire was developed to assess two kinds of OA pain[6] on the basis of aspects of pain that were identified as distinct and important by focus groups of people with OA[7]. *Constant pain* was characterised as a continuous aching sensation, and *intermittent pain*, was described as being severe but transient. This questionnaire was the first to assess these different types of OA pain, and it has been proposed that a total score may be a useful measure of overall pain severity in OA[6].

There are eleven items on the ICOAP forming two subscales considering both pain intensity and the effect of pain on quality of life[6]. Five items address constant pain and the remaining six items deal with intermittent pain. Items are responded to using a five-point scale. Ten items are phrased to assess the intensity of pain (e.g., *How intense has your constant knee pain been?*). The response options for these items are 0 (Not at all), 1 (Mildly), 2 (Moderately), 3 (Severely) or 4 (Extremely). In contrast, item 7 asks patients about the frequency of their pain (*How frequently has this knee pain that*

* Address correspondence and reprint requests to: B.J. Moreton, B26/27 International House, Institute of Work, Health and Organisations, Jubilee Campus, University of Nottingham, Nottingham NG8 1BB, UK. Tel: 44-115-846-6545; Fax: 44-115-846-6625.
E-mail addresses: bryan.moreton@nottingham.ac.uk (B.J. Moreton), maggie.wheeler@nottingham.ac.uk (M. Wheeler), david.walsh@nottingham.ac.uk (D.A. Walsh), nadina.lincoln@nottingham.ac.uk (N.B. Lincoln).

comes and goes occurred?). The response options for this item are 0 (Never), 1 (Rarely), 2 (Sometimes), 3 (Often) or 4 (Very often). The questionnaire can be administered by interview or self-completed[8].

The ICOAP questionnaire has shown good internal consistency, test–retest reliability and construct validity[6] when compared to scores on other questionnaires such as the Western Ontario and McMasters Universities Osteoarthritis (WOMAC) Index pain subscale[9]. It has also been successfully applied with other cultures[10–12] and is starting to be used as a measure of outcome in medical intervention studies[11,13,14]. The ICOAP is considered an important pain indexes by Osteoarthritis Research Society International (OARSI)[15], has been recommended by OA experts[16], and is included in several protocols for future trials[17,18]. However, there have been concerns raised over the dimensional structure of the questionnaire. Hawker *et al.*[6] conducted a principal components analysis of the ICOAP using a varimax rotation and found factorial complexity. Three components were extracted that accounted for 81.7% of the variance, but several items loaded onto more than one factor. These findings were supported in a secondary analysis using a promax rotation. The authors of the questionnaire therefore suggested using the total score rather than the subscales, but further testing was recommended[6].

The ICOAP hasn't been evaluated using Rasch analysis[19]. Rasch is a modern psychometric technique, appropriate for newly developed and existing questionnaires[5,20], which allows an examination of many critical measurement issues. It is important to allow researchers to ensure that different subgroups of participants (e.g., males and females) respond in similar ways given equivalent levels of pain[5] and to address concerns[6] about whether the ICOAP questionnaire displays unidimensionality[21]. Given fit to the model, it permits a transformation of ordinal data to interval-level measurement assisting with the calculation of mathematic operations such as change scores[22]. The purpose of the current study was therefore to examine fit between data obtained from the ICOAP and the Rasch model.

Methods

Participants and setting

Participants were recruited who had OA of the knee, defined and scored radiographically using Kellgren–Lawrence criteria[23], and reported accompanying pain on most days for at least the past month. Exclusion criteria were another rheumatic disease (e.g., Rheumatoid Arthritis, Gout and Psoriatic Arthritis), joint surgery within the 3 months prior to participation and an inability to speak or understand English.

Participants were identified from three sources: (1) a group that had taken part in a previous community-based study of knee OA[24], (2) Rheumatology and Orthopaedic clinics from Nottingham University Hospitals NHS Trust and (3) pre-operative assessment clinics from Sherwood Forest Hospitals NHS Trust. Potential participants were sent an invitation to the study, which was signed by a healthcare professional responsible for their care (e.g., surgeon). Those who agreed to participate were asked to complete a questionnaire set including measures of pain, anxiety, depression, fatigue, self-efficacy, acceptance, coping, beliefs, helplessness and quality of life. Only data from the ICOAP questionnaire and the Bodily Pain subscale of the RAND SF-36[25] are reported in the current article. The Bodily Pain subscale is composed of two items assessing pain over the past 4 weeks and the effect it has had on participants' ability to work. The items have a five and six point response format, respectively, and a total score is calculated from selected options whereby higher values indicate less pain. Rasch-transformed scores from the questionnaires were correlated using Pearson's coefficient to provide

an indication of external construct validity of the ICOAP. As participants were provided with a lengthy set of questionnaires they were advised to complete as many as they felt able to and the order of presentation was randomised into four sequences to minimise order effects. The questionnaires were completed at home and returned to the researchers by pre-paid envelope. Non-respondents were sent one reminder letter after 3 weeks as long as they were still able to participate.

Informed consent was obtained from all participants and the research was approved by Nottingham Research Ethics Committee one.

Rasch analysis

Three separate analyses were carried out on the Constant subscale, Intermittent subscale, and the Total scale. RUMM2020[26] was used to conduct the analyses.

Rasch analysis assesses fit between data obtained on a scale and the predictions of the Rasch model[22]. This is undertaken because the model shows us what to expect if a scale meets the axioms underlying additive conjoint measurement[5].

To help decide whether the rating scale[27] or the partial credit formulation[28] of the Rasch model was most appropriate, a likelihood ratio test was performed for each analysis. If the test was not significant (i.e., $P > 0.05$), then the rating scale version can be adopted; otherwise the partial credit version should be used. The tests were significant for all but the Intermittent subscale and so the partial credit formulation was used for consistency. However, both versions of the model resulted in similar conclusions. Individual items were inspected to see whether there was evidence of disordered response thresholds[5,22,29]. When this was observed, the item was rescored (i.e., collapsing appropriate adjacent response options).

Mean and standard deviation fit residuals were calculated for the items and the persons. These values were transformed to estimate a z-score representing standardised normal distribution and so, given good fit, the means should be close to 0 and the standard deviations about 1[22,29,30]. RUMM2020 creates groups, called class intervals, on the basis of the level of examined trait. An item-trait interaction chi-squared was used to test whether the hierarchical arrangement of the items was invariant across the class intervals. A significant *P*-value at the 0.05 level, with a Bonferroni adjustment for the number of items, signified that the item orderings differed across trait[22,30]. The Person Separation Index (PSI) was examined for each analysis to provide an indication of internal consistency. A value of 0.7 and above indicated acceptable internal consistency[29].

Each individual item and person was examined for misfit. For items, chi-squared and Analysis of Variance (ANOVA) fit statistics were calculated with a Bonferroni correction. Fit residuals were also examined for items and persons. Values above $+2.5$ or below -2.5 were considered to be misfitting the model[22].

Differential Item Functioning (DIF) was explored for gender (males and females) and age (<64 years, $64–71$ years and >71 years). When an item displays DIF it means that different subgroups produced significantly different responses despite having equivalent levels of trait[5,22]. ANOVA with a Bonferroni Correction was applied to explore DIF. Local independence of the items was examined in two ways. First, response dependencies between items were identified from the residual correlation matrix. A positive correlation of 0.3 or more was considered to be indicative of response dependency[31]. Second, each scale was tested for a breach of unidimensionality[5,21,32]. Principal components analysis was performed on the residuals and used to identify two subsets of items; those loading positively and negatively on the first

Table I
Participant characteristics

	n	%
Gender		
Females	91	52
Duration of OA*		
0–5 years	55	40
6–10 years	43	31
More than 10 years	39	28
Community group	99	57
Nottingham University Hospitals NHS Trust clinics	58	33
Sherwood Forest Hospitals NHS Foundation Trust clinics	18	10
	Median	**IQR†**
Age	66	60–73
Kellgren–Lawrence (tibiofemoral)	3	2–3
Kellgren–Lawrence (patellofemoral)	2	2–3
ICOAP constant	45	30–65
ICOAP intermittent	50	38–67
ICOAP total	48	34–65
RAND SF-36 Bodily Pain	45	23–58

* Length of OA was based on participant estimation. Some estimates were reported as uncertain and some participants did not provide a response. Percentages are based on the total responses available.

† IQR = inter-quartile range, 25th and 75th percentiles provided.

component. Person estimates were then calculated for each subset and then a series of independent t-tests were carried out to see whether the subsets produced significantly different estimates. Assuming that both subsets were measuring the same unidimensional construct no more than 5% of these t-tests should be significant at a 0.05 level. A binomial confidence interval (CI) was applied for cases that were more than 5%[29].

Results

Of the 474 people invited to take part, 175 provided data for analysis (37% response rate). Responses with three or more missing items were not included as recommended by the ICOAP user guide. The study sample had approximately 50% females and 50% males with a median age of 66 years (see Table I). About half of the participants came from the community group and half from hospital clinics. Gender and age information was available from 174 and 283, respectively, of the 299 people that were invited to take part but not included in the analysis. There were 89 females (51%) and the median age was 69 (inter-quartile range = 61–76). This suggests that the study sample was representative of the total group.

Constant subscale

All items exhibited ordered response thresholds. The summary item–person interaction statistics suggested misfit between data and

the model (see Table II). The standard deviation item fit residual (1.97) was high indicating that there was likely some misfit at an individual item level. Items 2 (*How much has your constant knee pain affected your sleep? Fit Residual = 3.49; χ^2 = 8.49, df = 2, P = 0.01; F = 3.85, df = 2, P = 0.02*) and 3 (*How much has your constant knee pain affected your overall quality of life? Fit Residual = −1.26; χ^2 = 10.12, df = 2, P = 0.006; F = 8.37, df = 2, P = 0.0003*) exhibited misfit; item 2 had a high positive fit residual indicating under discrimination. There was no evidence of DIF or response dependency. Principal components analysis of the residuals identified items that positively (items 1 and 2) and negatively (items 3, 4 and 5) loaded on the first component. Twenty-two out of 167 t-tests were significant, which represented 13.17% (Binomial CI: 9.90–16.50%) of the total tests.

To improve fit to the model item 2 was removed. This item was chosen for two reasons. First, like item 3, it exhibited misfit to the model. Second, Hawker et al.[6] found that item 2 (and item 8 – the Intermittent equivalent) loaded onto a separate factor indicating that it might be measuring sleep disorders in general as well as the effect of pain on sleep. Following removal of item 2 satisfactory fit to the model was achieved (see Table II). There were no misfitting items or persons, no evidence of response dependency or DIF, and the subscale passed the test of unidimensionality (5.42%; Binomial CI: 2.10–8.70%). Figure 1(a) shows the person–item threshold distribution and confirms that the subscale was well targeted. Only 5% of participants produced floor or ceiling effects.

Intermittent subscale

All item thresholds were ordered, but the summary fit statistics indicated some misfit between data and the model (see Table II). Item 9 (*How much has your knee pain that comes and goes affected your overall quality of life?*) had a significant F-statistic (*Fit Residual = −2.20; χ^2 = 5.81, df = 2, P = 0.05; F = 5.19, df = 2, P = 0.007*). Several participants also presented with high fit residuals. There was no evidence of response dependency, but item 8 (*How much has your knee pain that comes and goes affected your sleep?*) had uniform DIF for age. Specifically, participants aged 72 years and above produced lower expected values than the remaining participants across all three class intervals. Principal components analysis of the residuals identified items that loaded positively (items 9, 10 and 11) and negatively (items 6, 7 and 8) onto the first component. The Intermittent subscale marginally failed the test of unidimensionality (8.82%; Binomial CI: 5.50–12.10%).

The fit statistics may have been adversely affected by a small group of participants who responded in an unexpected way[22]. In total, 7% of participants produced fit residuals outside of the acceptable range. There was no obvious bias in this group towards sex

Table II
Summary fit statistics for the Constant subscale, Intermittent subscale and Total scale

| Rasch Analysis	Number of items	Item fit residual (Mean)	Item fit residual (SD)	Person fit residual (Mean)	Person fit residual (SD)	χ^2 (df)	P-value	PSI	Percentage of significant t-tests (CI)
Constant (initial)	5	0.03	1.97	−0.44	1.13	20.82 (10)	P = 0.02	0.92	13.17% (9.90–16.50%)
Constant (remove 2)	4	−0.08	1.23	−0.43	0.89	7.44 (8)	P = 0.49	0.92	5.42% (2.10–8.70%)
Intermittent (initial)	6	0.06	2.13	−0.47	1.26	22.25 (12)	P = 0.03	0.90	8.82% (5.50–12.10%)
Intermittent (remove 9)	5	0.23	1.47	−0.42	1.13	11.88 (10)	P = 0.29	0.87	7.06% (3.80–10.30%)
Intermittent (remove 8)	4	0.17	1.59	−0.47	1.01	5.64 (8)	P = 0.69	0.85	7.14% (3.80–10.40%)
Total scale (initial)	11	0.08	1.89	−0.72	1.84	37.47 (22)	P = 0.02	0.94	9.25% (6.00–12.50%)
Total scale (rescore 7)	11	0.09	1.88	−0.71	1.84	35.05 (22)	P = 0.04	0.94	9.25% (6.00–12.50%)
Total scale (remove 7 & 6)	9	−0.01	1.51	−0.79	1.78	27.72 (18)	P = 0.07	0.93	15.61% (12.40–18.90%)
Total scale (remove 8 & 2)	7	−0.18	1.26	−0.83	1.51	21.42 (14)	P = 0.09	0.93	6.36% (3.10–9.60%)
Ideal value		*0.00*	*1.00*	*0.00*	*1.00*		*>0.05*	*>0.70*	*<5%*
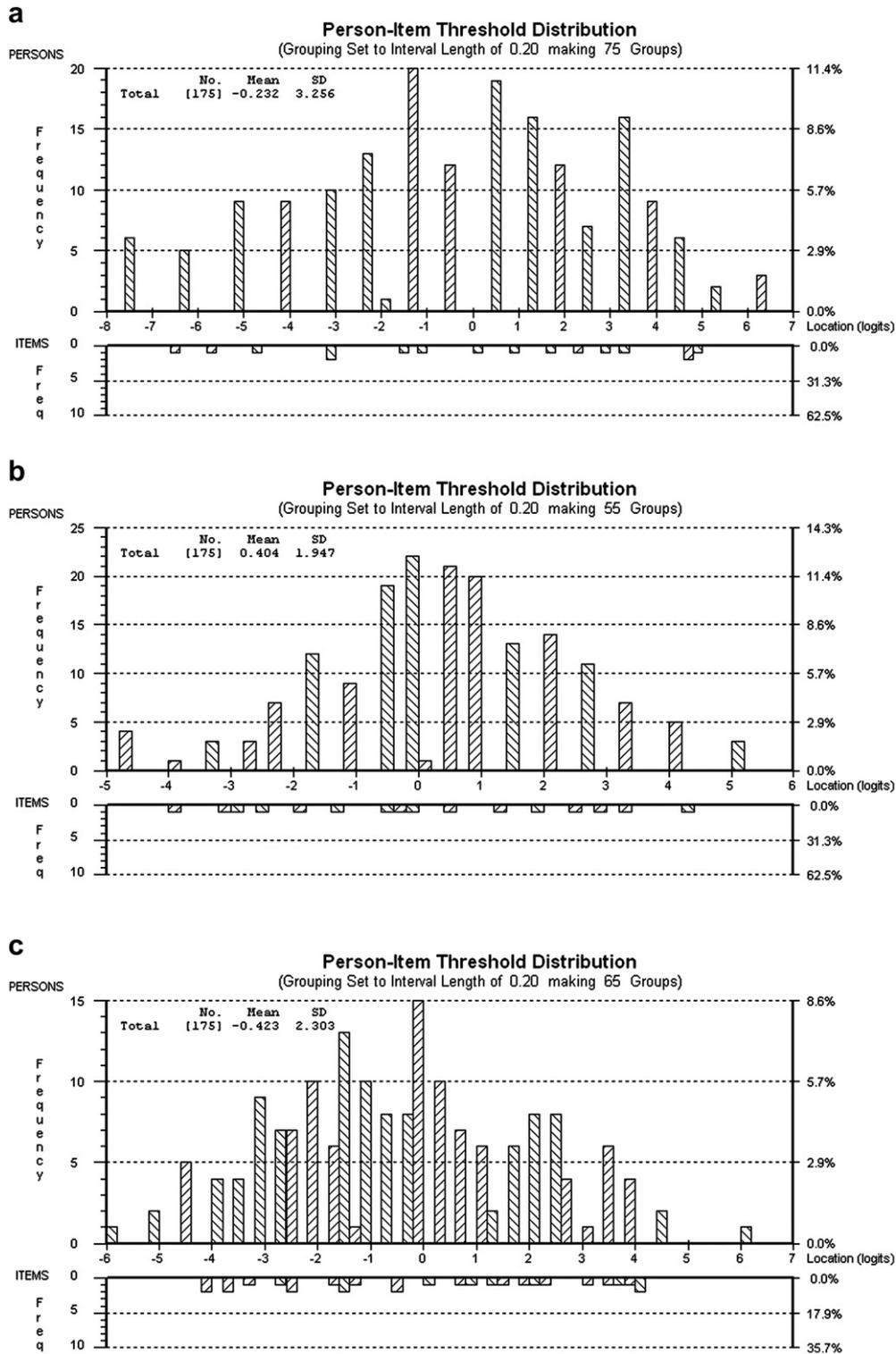
Fig. 1. Person–item threshold distributions for the Constant subscale (a), Intermittent subscale (b) and Total scale (c). The top part of the graphs show the distribution of persons along the Rasch-transformed scale of constant, intermittent and total pain, respectively. The bottom part shows the distribution of items. Ideally, given good targeting, the two distributions should mirror one another such that the full range of trait expressed by participants is captured by the items on a scale.

(54% female), nor towards a particular age group (<64 years = 31%; 64–71 years = 38%; >71 years = 31%). The majority of participants had high negative fit residuals (85%), which indicates that their responses were too deterministic (i.e., too similar to a Guttman pattern) for the Rasch model. However, the summary fit statistics

displayed in Table II suggest that the overall fit for persons was relatively good (e.g., SD < 1.4). It is possible that floor or ceiling effects increased the chance of obtaining a negative fit residual, but only 3% of participants produced them. Therefore, a technique proposed by Linacre[33] was used to examine whether it was necessary to remove

the misfitting participants. Persons estimates before and after removal of the misfitting participants were plotted against each other. Inspection of the plot confirmed that the effect of removing the participants was minimal and so they were retained.

Item 9 had a significant ANOVA test and so was considered for deletion. Removal of this item resulted in improved fit as shown in Table II. Item 8 continued to exhibit DIF for age. As the DIF was uniform, an attempt was made to 'split the DIF'. This is where the item is split according to the subgroups that produced different scores – in this case 71 years and below and 72 years and above. However, this resulted in an increase in the item–trait interaction χ^2 (11.88–18.83), which suggests reduced fit. To examine whether this item was having a negative effect on the fit statistics it was removed, which improved the overall fit (see Table II). The revised 4-item Intermittent subscale had no misfitting items or persons, no response dependency or DIF and was unidimensional (7.14%; Binomial CI: 3.80–10.40%). Following these changes item 7 exhibited marginally disordered thresholds but rescoring didn't improve the fit statistics and so the original scoring was retained. The person–item threshold distribution showed reasonable targeting [see Fig. 1(b)]. Four percent of participants were at floor or ceiling levels.

Total scale

Item 7 (*How frequently has this knee pain that comes and goes occurred?*) had disordered thresholds and so was rescored by collapsing response options 0 and 1. The summary fit statistics suggested significant misfit for both items and persons (see Table II). Item 7 misfits the model (*Fit Residual* = 3.67; χ^2 = 8.63, *df* = 2, *P* = 0.01; *F* = 3.85, *df* = 2, *P* = 0.02) and 18% of participants presented with high fit residuals. There was evidence of response dependency: items 2 and 8 (0.62), items 4 and 5 (0.40), items 9 and 10 (0.33) and items 10 and 11 (0.34). Notice that with the exception of items 2 and 8 the response dependency was limited to items within the same subscales. Items 8 [see Fig. 2(a)] and 9 [see Fig. 2(b)] exhibited uniform and non-uniform DIF for age, respectively. The Total scale failed the test of unidimensionality with 16 out of 173 *t*-tests significant (9.25%; Binomial CI: 6.60–12.50%). The subsets of items formed from the principal components analysis of the residuals

Fig. 2. DIF across age groups for items 8 (a) and 9 (b) of the Total scale.

were potentially revealing of the underlying cause of the multidimensionality. Specifically, items 1, 2, 3, 4, 5 and 8 (with the exception of 8, all Constant items) positively loaded onto the first component and items 6, 7, 9, 10 and 11 (all Intermittent items) negatively loaded.

To improve fit to the model, items 7, 6 (*How intense has your most severe knee pain that comes and goes been? Fit Residual* = 2.96; χ^2 = 11.78, *df* = 2, *P* = 0.003; *F* = 4.60, *df* = 2, *P* = 0.01), 8 (*Fit Residual* = 2.65; χ^2 = 5.79, *df* = 2, *P* = 0.06; *F* = 2.45, *df* = 2, *P* = 0.09) and 2 (*Fit Residual* = 3.93; χ^2 = 7.68, *df* = 2, *P* = 0.02; *F* = 3.21, *df* = 2, *P* = 0.04) were removed one-by-one due to misfit (misfit for items 6, 8 and 2 became apparent after removing the preceding item). As Table II shows, this improved the fit statistics. There were no more misfitting items, no response dependency or DIF and the scale was unidimensional (6.36%; Binomial CI: 3.10–9.60%). The remaining Constant (1, 3, 4 and 5) and Intermittent (9, 10 and 11) items continued to load in opposite directions on the first component. The revised 7-item Total scale was relatively well targeted [see Fig. 1(c)], but unlike the analysis of the Intermittent subscale, the number of participants misfitting the model was still high (11%) following the changes. This was particularly reflected in the mean person fit residual (see Table II). This wasn't attributable to floor or ceiling effects, which were low (1%). Alternative analysis plans were carried out (i.e., removing misfitting persons before altering the scale, subtesting for response dependency and removing items 2 and 8 first), but they resulted in similar conclusions.

Correlation analysis

Data from the Bodily Pain subscale of the RAND SF-36 were Rasch analysed and suitable fit was achieved after minimal changes (rescoring one item). Pearson correlation coefficients were used to examine the relationships between the interval scores produced from the questionnaires. The Constant ($r = -0.69$, $P < 0.0001$) and Intermittent ($r = -0.65$, $P < 0.0001$) subscales significantly and negatively correlated with the Bodily Pain subscale suggesting that both measure pain.

Discussion

The current study explored the fit between data obtained from the ICOAP questionnaire and the Rasch model. The Constant subscale, Intermittent subscale and Total scale were analysed separately. Item 2 was removed from the Constant subscale due to misfit and multidimensionality. This resulted in adequate fit to the model. Items 8 and 9 were deleted from the Intermittent subscale because of DIF and misfit, respectively, which improved the fit statistics. The Total scale exhibited evidence of response dependency and multidimensionality, which violates the assumption of local independence[5,22,34,35]. There was also evidence of item and person misfit and DIF. Four items were removed from the analysis (items 2, 6, 7, and 8), which resolved the response dependency and multidimensionality and improved the fit. However, a number of participants continued to exhibit misfit after the scale was altered.

Hawker *et al.*[6] observed factorial complexity when exploring the ICOAP with principal components analysis. The current study showed that the Constant and Intermittent subscales meet the requirements of the Rasch model once a few items have been removed. Principal components analysis of the residuals for the Total scale showed multidimensionality with items from the Constant and Intermittent subscales mostly loading in opposite directions. This suggests that the ICOAP is measuring both types of pain separately. Although removal of a few misfitting items from the Total scale resulted in an unidimensional measure supporting the findings of Hawker *et al.*, the Constant and Intermittent items continued to load in opposite directions even after changes were
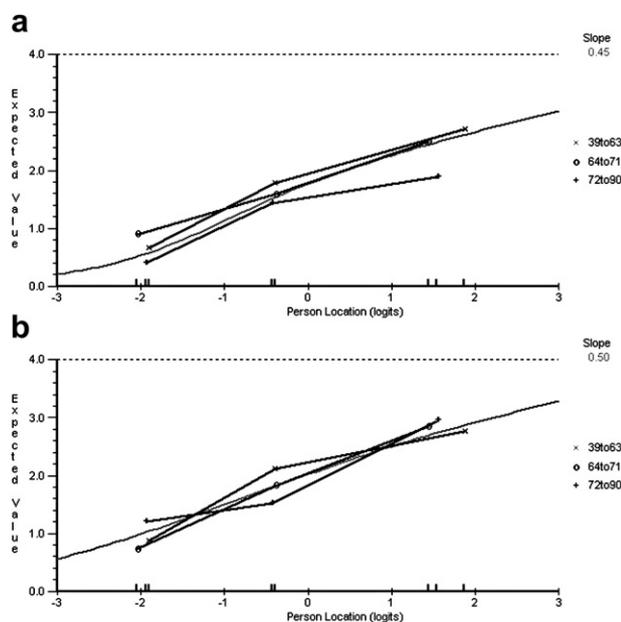
made. Considering these findings it is suggested that the ICOAP is used as a measure of two different types of pain, as originally designed, rather than a general measure of pain.

Response dependency was observed between several items in the Total scale, but with the exception of items 2 and 8 it only occurred between items within the same subscales. This is important because the items ask participants about constant and intermittent pain in the same way (e.g., *Items 3/9 — how much has your constant knee pain/knee pain that comes and goes affected your overall quality of life*), which means that the participants were able to separate their sensations of constant and intermittent pain. If they were not able to do this, then their responses to items on the Constant subscale would have an impact on their responses to items on the Intermittent subscale creating response dependency.

These findings highlight the multidimensional nature of pain experience. It is well recognised that numerous factors need to be considered when assessing the mechanisms of pain such as physical (e.g., function), emotional (e.g., depression), social (e.g., family support) and cognitive (e.g., coping strategies)[36]. This study has shown that pain can be subdivided according to the periodicity of symptoms, which is well-captured by the ICOAP questionnaire.

The study has some limitations that need to be considered. The Intermittent subscale initially had misfitting participants. However, this was resolved by alterations to the scale. In contrast, a few participants misfit the Total scale even after changes were made. This raises concerns over the external validity of the scale[22]. Most of the misfitting participants (79%) had high negative fit residuals, which means that their responses were too deterministic[37]. These participants tended to select mostly the same response option for each item leading to less variation in their answers. This may have been due to them being given more than one questionnaire to complete leading to fatigue and less thought about each item[38].

The sample size of the current study was slightly small. However, it was sufficient to have 99% confidence that the item calibrations were within ±0.5 logits[39]. It is useful to have more data, but it is not always feasible. Sample size requirements in Rasch analysis are in part governed by targeting. The results of this study have shown that the ICOAP is relatively well targeted with knee OA participants. The response rate was a little low in the study, which raises concerns over the generalisability of the findings. However, comparable response rates have been reported in similar studies[30] and the available demographic information from those who did not respond suggests that they were similar in terms of gender and age. Nevertheless it is recommended that further Rasch analyses are conducted to corroborate these results.

There are two versions of the ICOAP; one for knee OA and the other for hip OA. The current study only used the knee OA version because only these patients were included. Further studies examining the fit between the hip version and the Rasch model would be useful. It would be interesting to investigate whether knee and hip OA patients produce different ICOAP scores when their estimated levels of pain are equivalent. Thus a study including both patient types and an examination of DIF would be particularly desirable.

In conclusion, the Constant and Intermittent subscales of the ICOAP fit the Rasch model following removal of a few items. The Total scale can be adjusted to improve its fit to the model, but significant changes were required and a number of participants misfit raising concerns over its external validity. It is recommended that the Constant and Intermittent subscales are used rather than the Total scale. Clinicians and researchers wishing to use parametric statistics with these subscales may use the conversion values provided in the Appendix.

Contributions

Role of funding source

Conflict of interests

Acknowledgements

Appendix

Items 1, 3, 4, and 5 from the Constant Pain subscale should be summed to provide a score from 0 to 16. Items 6, 7, 10 and 11 from the Intermittent Pain subscale should be summed to provide a score from 0 to 16. The raw score can then be converted to an interval scale using the corresponding 'Rasch score' value.

Subscale	Raw score	Rasch score	Standard error
Constant pain	0	−7.48	1.47
	1	−6.29	1.19
	2	−5.17	1.07
	3	−4.10	1.01
	4	−3.14	0.98
	5	−2.22	0.96
	6	−1.32	0.94
	7	−0.44	0.93
	8	0.42	0.91
	9	1.22	0.87
	10	1.94	0.83
	11	2.60	0.82
	12	3.23	0.82
	13	3.88	0.84
	14	4.58	0.90
	15	5.39	1.05
	16	6.35	1.43
Intermittent pain	0	−4.77	1.42
	1	−3.94	0.97
	2	−3.27	0.81
	3	−2.73	0.76
	4	−2.20	0.75
	5	−1.65	0.74
	6	−1.10	0.73
	7	−0.57	0.73
	8	−0.06	0.72
	9	0.46	0.73
	10	1.00	0.74
	11	1.55	0.75
	12	2.11	0.77
	13	2.70	0.80
	14	3.34	0.88
	15	4.16	1.05
	16	5.20	1.44

References

1. Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. Ann Rheum Dis 2001;60:91–7.

2. Dieppe PA, Lohmander LS. Pathogenesis and management of pain in osteoarthritis. Lancet 2005;365:965–73.

3. Rosemann T, Laux G, Szecsenyi J, Wensing M, Grol R. Pain and osteoarthritis in primary care: factors associated with pain perception in a sample of 1,021 patients. Pain Med 2008;9:903–10.

4. Katz JN, Melzack R. Measurement of pain. Surg Clin North Am 1999;79:231–52.

5. Tennant A, Conaghan P. 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 2007;57:1358–62.

6. Hawker GA, Davis AM, French MR, Cibere J, Jodan JM, March L, et al. Development and preliminary psychometric testing of a new OA pain measure – an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16:409–14.

7. Hawker GA, Stewart L, French MR, Cibere J, Jodan JM, March L, et al. Understanding the pain experience in hip and knee osteoarthritis – an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16:415–22.

8. Hawker GA. Measure of intermittent and constant osteoarthritis pain: ICOAP. In: OARSI User Guide, 2010.

9. Bellamy N. Pain assessment in osteoarthritis: experience with the WOMAC osteoarthritis index. Semin Arthritis Rheum 1989;18:14–7.

10. Gonçalves RS, Cabri J, Pinheiro JP, Ferreira PL, Gil J. Cross-cultural adaptation and validation of the Portuguese version of the intermittent and constant osteoarthritis pain (ICOAP) measure for the knee. Osteoarthritis Cartilage 2010;18:1058–61.

11. Kessler S, Grammozis A, Günther KP, Kirschner S. Testtheoretische Überprüfung der deutschen Version des Intermittent and Constant Osteoarthritis Pain Score (ICOAP) – ein Fragebogen zur Schmerzerfassung bei Patienten mit Gonarthrose. Z Orthop Unfallchir 2011;149:22–6.

12. Maillefert JF, Kloppenburg M, Fernandes L, Punzi L, Günther KP, Martin Mola E, et al. Multi-language translation and cross-cultural adaption of the OARSI/OMERACT measure of intermittent and constant osteoarthritis pain (ICOAP). Osteoarthritis Cartilage 2009;17:1293–6.

13. Davis AM, Lohmander LS, Wong R, Venkataramanan V, Hawker GA. Evaluating the responsiveness of the ICOAP following hip or knee replacement. Osteoarthritis Cartilage 2010;18:1043–5.

14. Bond M, Davis AM, Lohmander LS, Hawker GA. Responsiveness of the OARSI–OMERACT osteoarthritis pain and function measures. Osteoarthritis Cartilage 2012;20:541–7.

15. OARSI. OARSI initiatives: pain indexes, http://www.oarsi.org/index2.cfm?section=OARSI_Initiatives&content=Pain_Radiological_Indexes.

16. Wright T, Goldring S. Reaching consensus and highlighting future directions for research: the osteoarthitis summit breakout sessions. HSS J 2012;8:80–3.

17. van Es PP, Luijsterburg PAJ, Dekker J, Koopmanschap MA, Bohnen AM, Verhaar JAN, et al. Cost-effectiveness of exercise therapy versus general practitioner care for osteoarthritis of the hip: design of a randomised clinical trial. BMC Musculoskelet Disord 2011;12:232.

18. Verkleij SPJ, Luijsterburg PAJ, Koes BW, Bohnen AM, Bierma-Zeinstra SMA. Effectiveness of diclofenac versus acetaminophen in primary care patients with knee osteoarthritis: [NTR1485], DIPA-trial: design of a randomized clinical trial. BMC Musculoskelet Disord 2010;11:7.

19. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago; 1960.

20. Ko Y, Lo NN, Yeo SJ, Yang KY, Yeo W, Chong HC, et al. Rasch analysis of the Oxford Knee Score. Osteoarthritis Cartilage 2009;17:1163–9.

21. Tennant A, Pallant JF. Unidimensionality matters! (A tale of two smiths?). Rasch Meas Trans 2006;20:1048–51.

22. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 2007;46:1–18.

23. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. Ann Rheum Dis 1957;16:494–501.

24. Doherty M, Hawkey C, Goulder M, Gibb I, Hill N, Aspley S, et al. A randomised controlled trial of ibuprofen, paracetamol or a combination tablet of ibuprofen/paracetamol in community-derived people with knee pain. Ann Rheum Dis 2011;70:1534–41.

25. Hays RD, Prince-Embury S, Chen H. RAND-36 Health Status Inventory. San Antonio, TX: The Psychological Corporation; 1998.

26. Andrich D, Lyne A, Sheridan B, Luo G. RUMM2020. Perth: RUMM Laboratory; 2003.

27. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978;43:561–73.

28. Masters GN, Wright BD. The partial credit model. In: van der Linden WJ, Hambleton RK, Eds. Handbook of Modern Item Response Theory. New York: Springer; 1997.

29. Shea TL, Tennant A, Pallant JF. Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). BMC Psychiatry 2009:9.

30. Pallant JF, Keenan A-M, Misajon RA, Conaghan PG, Tennant A. Measuring the impact and distress of osteoarthritis from the patients' perspective. Health Qual Life Outcomes 2009;7:37.

31. Ramp M, Khan F, Misajon RA, Pallant JF. Rasch analysis of the Multiple Sclerosis Impact Scale (MSIS-29). Health Qual Life Outcomes 2009;7:58.

32. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3:205–31.

33. Linacre JM. When to stop removing items and persons in Rasch misfit analysis? Rasch Meas Trans 2010;23:1241.

34. Baghaei P. Local dependency and Rasch measures. Rasch Meas Trans 2008;21:1105–6.

35. Wright BD. Local dependency, correlations and principal components. Rasch Meas Trans 1996;10:509–11.

36. Keefe FJ, Somers TJ. Psychological approaches to understanding and treating arthritis pain. Nat Rev Rheumatol 2010;6:210–6.

37. Curtis DD. Person misfit in attitude surveys: influences, impacts and implications. Intern Educ J 2004;5:125–44.

38. Younger J, McCue R, Mackey S. Pain outcomes: a brief review of instruments and techniques. Curr Pain Headache Rep 2009;13:39–43.

39. Linacre JM. Sample size and item calibration stability. Rasch Meas Trans 1994;7:328.