ORIGINAL REPORT

CAN THE ICF OSTEOARTHRITIS CORE SET REPRESENT A FUTURE CLINICAL TOOL IN MEASURING FUNCTIONING IN PERSONS WITH OSTEOARTHRITIS UNDERGOING HIP AND KNEE JOINT REPLACEMENT?

Maria Jenelyn Alviar, MD, MSc[1], John Olver, MBBS, MD, FAFRM[2], Julie F. Pallant, PhD[3], Caroline Brand, MBBS, MPH, FRACP[4], Richard de Steiger, MBBS, FRACS[5], Marinis Pirpiris, MBBS, PhD, FRACS[6], Andrew Bucknill, MBBS, MSc, FRACS[6] and Fary Khan, MBBS, MD, FAFRM[4]

From the [1]The University of Melbourne-Royal Melbourne Hospital, [2]Monash University-Epworth Rehabilitation, [3]Rural Health Academic Centre, The University of Melbourne, [4]The University of Melbourne-Royal Melbourne Hospital and Monash University, [5]Epworth Healthcare and The University of Melbourne and [6]Royal Melbourne Hospital, Melbourne, Australia

Objective: To determine the dimensionality, reliability, model fit, adequacy of the qualifier levels, response patterns across different factors, and targeting of the International Classification of Functioning, Disability and Health (ICF) osteoarthritis core set categories in people with osteoarthritis undergoing hip and knee arthroplasty.

Methods: The osteoarthritis core set was rated in 316 persons with osteoarthritis who were either in the pre-operative or within one year post-operative stage. Rasch analyses were performed using the RUMM 2030 program.

Results: Twelve of the 13 body functions categories and 13 of the 19 activity and participation categories had good model fit. The qualifiers displayed disordered thresholds necessitating rescoring. There was uneven spread of ICF categories across the full range of the patients' scores indicating off-targeting. Subtest analysis of the reduced ICF categories of body functions and activity and participation showed that the two components could be integrated to form one measure.

Conclusion: The results suggest that it is possible to measure functioning using a unidimensional construct based on ICF osteoarthritis core set categories of body functions and activity and participation in this population. However, omission of some categories and reduction in qualifier levels are necessary. Further studies are needed to determine whether better targeting is achieved, particularly during the pre-operative and during the sub-acute care period.

Key words: ICF; outcome assessment; arthroplasty; joint replacement; Rasch measurement; osteoarthritis.

J Rehabil Med 2012; 44: 955–961

Correspondence address: Maria Jenelyn Alviar, The University of Melbourne (Parkville Campus), Victoria 3010, Australia. E-mail: jen.alviar@gmail.com or m.alviar@pgrad.unimelb.edu.au

INTRODUCTION

Functioning is one of the most crucial outcomes in the rehabilitation of persons with osteoarthritis (OA) undergoing arthroplasty.

As OA is a chronic degenerative disease with no cure and elective joint replacement surgeries are becoming more frequent, functioning is an important concern among those afflicted. Information on functioning is essential over the course of their clinical journey from the time that they are seen in the out-patient clinic, in the surgical ward, during rehabilitation and then back in the community. However, the varying taxonomy of functioning has left the concept open to various interpretations leading to confusion (1). At present, a wide variety of measures are used in arthroplasty but there is little consensus on which ones to use (2). Furthermore, patient-reported outcome measures applied in arthroplasty rehabilitation do not adequately address relevant areas of activity, participation and environment (3).

In recent years, the International Classification of Functioning, Disability and Health (ICF) model has provided a unifying definition for functioning (4). The huge size of the ICF, however, can appear daunting for use in a real world setting. Thus, short lists of relevant categories for specific conditions have been grouped in core sets (5). Describing and assessing functioning using the ICF entails rating categories with the qualifier on an ordinal 5-point scale (4). Qualifier ratings across categories of a core set yield a categorical profile which represents the functional state of an individual (6). This may then be used as a guide in rehabilitation goal-setting, intervention planning, and follow-up. With this perspective, the ICF and its core sets have the potential to be practical, applicable, and useful in the assessment of functioning in arthroplasty rehabilitation in the busy clinical setting. From a select group of ICF categories, such as those in the OA core set, an ICF-based clinical measure that is reduced in form could provide summary scores representing indices of functioning. Such scores would be useful and practical for clinicians in assessing and monitoring the level of functioning of patients during the course of their care and thoughout their lives. In this way, a common language is also achieved for all disciplines involved in the care of this population. On a research level, this allows for comparisons and pooling of scores from various settings.

Studies have explored the possibility of constructing clinical measures of functioning across relevant ICF categories (core

sets) in some other health conditions (7, 8, 9, 10).Their findings provide support that the ICF, although a classification system, could be a starting point for future clinical instruments. So far, creation of an ICF-oriented clinical measure of functioning for persons with OA undergoing hip and knee joint replacement has not been undertaken. This study utilises Rasch analysis, which is becoming a prominent tool in rehabilitation medicine research as a measurement model on which assessment instruments for functioning can be developed (11).

The aim of the current study is to determine whether the OA core set can be utilized in the development of a clinical measure of functioning specifically for persons with OA undergoing hip and knee arthroplasty. By doing so, the study adds to the body of knowledge on the practicality, clinical utility, and applicability of the ICF in real world settings. Thus, this study specifically aims to determine the dimensionality, reliability, item fit, adequacy of the qualifier levels, item bias, and targeting of the ICF OA core set categories in a hip and knee joint replacement cohort using Rasch analysis.

METHODOLOGY

Design and Setting

The study used a cross-sectional design involving two tertiary hospitals (one public and one private) in Australia. The study was approved by the Human Research Ethics Committee in each involved hospital.

Participants

Adults with OA who were either waiting for hip or knee arthroplasty (pre-operative), or within one year of surgery (post-operative), and able to read and understand English were recruited to participate in the study.

Data collection

A rehabilitation physician and a nurse, who were both oriented and trained in the study procedures, obtained the consent, collected the data and scored the ICF comprehensive core set for OA based on an interview and all available clinical information from March 2010 to July 2011. The Kappa statistic between the two raters was 0.79 indicating substantial agreement. The OA comprehensive core set is composed of 55 categories organized into 4 components, namely, body structures, body functions, activity and participation and environmental factors. It consists of 6 categories from body structures, 13 from body functions, 19 from activities and participation, and 17 from environmental factors (12). Only the categories pertaining to the functioning dimension were considered in this study. The 5-point qualifier scale, ranging from 0–4, was used to assess the extent of the problem a participant might have in each of the ICF categories in the core set: none (0), mild (1), moderate (2), severe (3), complete (4) problem/difficulty (4). The qualifiers "8" (not specified) and "9" (not applicable) were also used.

Data analysis

Descriptive statistics were used to describe the characteristics of the study population.

Rasch analysis was applied using RUMM 2030 (13) to assess dimensionality, over-all fit of the model, individual item fit, adequacy of the response options, differential item functioning (DIF) across age, sex, and educational level, and targeting of the ICF OA core set. As the ICF OA core set has polytomous response options, a likelihood

ratio test was conducted to determine whether it was more appropriate to use the rating scale model or the partial credit model. Rasch analysis of the ICF OA core set was done in two stages. The first stage was to separately analyse each component of the dimension functioning (body structures, body functions, activity and participation). The second stage was to assess the suitability of integrating all the components of the OA core set as a measure of functioning. The qualifiers "8" and "9" were considered missing values. At the outset, based on descriptive statistics, categories with more than 10% missing data were excluded from the analysis except when the category was deemed clinically relevant.

Stage 1 Rasch. Chi square item-trait interaction statistics were applied to evaluate the overall fit of the model for the ICF categories of each component.In Rasch analysis, an outcome scale is tested to determine whether it satisfies a mathematical measurement model (14). A more detailed description of Rasch analysis procedures is described elsewhere (15, 16, 17). A significant probability indicates some degree of misfit between the data and the model (16, 17). A Bonferroni-corrected significance level is used to adjust for multiple comparisons (18). The item person interaction statistics for items and persons were used to assess item fit and person fit to the model. The standard deviation (SD) of the summary residual statistic for items and persons should be less than 1.5. Individual item fit residual values should be between –2.5 and +2.5 to indicate adequate model fit (17).

The suitability of the response format was evaluated by examining the category probability curves and threshold map for disordered response thresholds. Where disordered thresholds occurred, the response options were collapsed to improve overall fit to the model (17).

Items were examined for DIF with respect to age (dichotomised at 65 years), sex, and education (tertiary vs. non-tertiary) using analysis of variance for each category with a Bonferroni-adjusted α-level. Differential item functioning is a form of item bias signifying that different groups in the sample, despite similar levels of underlying trait, respond differently to an item (17).

Internal consistency reliability was assessed using the Person Separation Index (PSI). The PSI also evaluates the ability of the measure to discriminate among persons with different levels of the trait. Values range from 0 to 1 and minimum values of 0.7 and 0.85 indicate adequate reliability for group use and for individual use, respectively (16).

The targeting of the ICF categories and the persons' abilities was assessed by inspecting the item threshold map and by comparing the mean location score obtained by the persons with that of the value of zero set for the items. For a well-targeted measure, the mean location for the persons would be around the value of zero (16, 17).

Local dependency was assessed by examining the residual correlations for values of 0.3 and above (17). To assess dimensionality, principal components analysis (PCA) was performed on the residuals. Factor loadings on the first component were used to identify two different subsets of items (i.e. the positively and negatively correlated items). A series of *t*-tests were performed for each person in the sample comparing their scores on the two subsets. If more than 5% of the sample has statistically different scores on each of the subsets of items, this suggests that the items in the two subsets may be tapping different constructs and thus the scale may not be unidimensional (19).When the value of 5% is exceeded, a 95% confidence interval (CI) was computed for the observed number of significant tests. If the interval contains the value of 5%, the scale is considered unidimensional.

Stage 2 Rasch. After analysing the individual ICF components, modifications (rescoring items, deleting items) were made to ensure adequate fit to the Rasch model. The modified components were subjected to subtest analysis in RUMM2030 to evaluate the suitability of combining the components to form a total scale. The dimensionality was assessed with *t*-tests to check whether these components can form one single scale. If more than 5% of these tests were significant, then these components are measuring two different constructs and therefore could not be combined to form one scale.

RESULTS

The total number of eligible participants from the two institutions was 975. Twenty percent (199) were excluded due to language restriction. The remainder (776) were invited to participate through letters of invitations. Four hundred and sixty declined (47%) and one withdrew during the interview. All 316 participants were Australian residents with a mean age of 67 years and 59% were females. Forty-two percent had primary hip replacement and 34% had primary knee replacement. Almost half of them were in the 3–12 months post-operative stage, a third within 3 months post-operative stage, and the remainder were in the pre-operative stage (Table I).

Rasch analysis

Prior to Rasch analysis, the likelihood ratio test was conducted and resulted in a significant test ($p < 0.001$), supporting the use of the partial credit model.

Body functions

Initial analysis of the 13 items (categories) assessing body functions revealed a significant item-trait interaction χ^2 statistic χ^2 ($p = 0.0016$) suggesting poor overall fit to the model (Table II, Analysis 1). Inspection of the threshold map revealed disordered thresholds for nearly all categories. This was resolved by collapsing response options by rescoring from 01234 to 01123 (Table II, Analysis 2). Categories *b134 sleep functions and b710 mobility of joint functions* had fit residual values outside +/–2.5. Deletion of *b710* resulted in adequate model fit with a non-significant χ^2 ($p = 0.029$), after using a Bonferroni-adjusted *p*-value (Table II, Analysis 3). The mean fit residuals and SDs for items and persons were within acceptable limits. The PSI

Table I. *Characteristics of the study population (*n = 316)

Age, years, mean (SD)	67 (10)
<65 years, *n* (%)	124 (39)
≥65 years, *n* (%)	192 (61)
Gender, *n* (%)	
Male	129 (41)
Female	187 (59)
Status, *n* (%)	
Single	30 (9)
Married	209 (66)
Widowed/divorced	72 (24)
Education, *n* (%)	
Non-tertiary	188 (59)
Tertiary	121 (38)
Occupation, *n* (%)	
Employed	90 (29)
Retired/unemployed	226 (71)
Duration of arthritis, years, mean (SD)	13 (11)
Type of joint replacement, *n* (%)	
Primary hip	131 (41)
Primary knee	110 (35)
Revision hip	3 (1)
Revision knee	4 (1)
Surgery status, *n* (%)	
Pre-op	68 (21)
Within 3 months post-op	91 (29)
>3–12 months post-op	157 (50)
Type of hospital, *n* (%)	
Public	108 (34)
Private	208 (66)
10-cm VAS, mean (SD)	3 (3)
FIM Motor, median (IQR)	89 (85–90)
WOMAC physical function, median (IQR)	5 (1–16)

SD: standard deviation; VAS, visual analog scale; FIM: Functional Independence Measure; IQR: interquartile range; WOMAC: Western Ontario McMaster University Osteoarthritis Index.

Table II. *Summary of results of the Rasch analyses (*n = 316)

Action	Analysis	Over-all model fit	Item fit residual Mean (SD)	Person fit residual Mean (SD)	PSI	% significant *t*-tests[a]
Body functions						
Original component	1	$\chi^2 = 70.26$ $p = 0.002$	–0.57 (1.44)	–0.25 (0.75)	0.67	3.51%
Rescoring of categories	2	$\chi^2 = 70.90$ $p = 0.001$	–0.33 (1.49)	–0.24 (0.82)	0.70	4.39%
Removal of b710	3	$\chi^2 = 53.67$ $p = 0.029$	–0.35 (1.14)	–0.26 (0.78)	0.66	7.69% [95% CI 4–12%]
Activity & Participation						
Removal of d770, d850, d470 from original component based on descriptive statistics	4	$\chi^2 = 21.69$ $p < 0.001$	–0.48 (2.25)	–0.21 (0.90)	0.70	3.57%
Rescoring of all items	5	$\chi^2 = 271.06$ $p < 0.001$	–0.48 (2.33)	–0.24 (0.93)	0.74	9.80% [95% CI 4–16%]
Removal of d445	6	$\chi^2 = 226.75$ $p < 0.001$	–0.43 (2.11)	–0.23 (0.90)	0.75	5.71% [95% CI 2–10%]
Removal of d440	7	$\chi^2 = 62.57$ $p = 0.021$	–0.13 (1.41)	–0.24 (0.90)	0.77	2.42%
Removal of d450	8	$\chi^2 = 48.04$ $p = 0.15$	–0.12 (1.00)	–0.24 (0.85)	0.73	4.17%
Subtest analysis						
Subtest	9	$\chi^2 = 15.57$ $p = 0.02$	–0.10 (0.58)	–0.36 (0.47)	0.70	1.71%

[a]CIs are only reported when % exceeds 5%. SD: standard deviation; PSI: person separation index; χ^2: chi square; *p*: probability.

was 0.66. No DIF was detected for age, sex, and educational level. Local dependency was checked and there were no correlations above 0.3.

Dimensionality testing, comparing the person estimates generated from the two subsets identified from PCA, indicated that 9 (7.69%) of the 117 *t*-tests had significant differences in the estimates. A 95% CI contained the value of 5% supporting unidimensionality (Table II, Analysis 3).

The left panel of the item map (Fig. 1) shows the distribution of participants along the Rasch calibrated scale and the right panel shows the difficulty level of items (categories). The easiest categories (closer to the bottom) were *b280 sensation of pain, b740 muscle endurance functions,* and *b770 gait pattern functions*. The most difficult categories were *b152 emotional functions, b760 control of voluntary movement functions,* and *b130 energy and drive functions*. The mean location value for persons was –3.18 (SD 1.43). There are gaps at the bottom of the map, where no items match the ability of persons and at the top, where no persons correspond to the items (Fig. 1).

Body structures

For body structures, the category *s799 structures related to movement, unspecified* had more than 50% missing values and was excluded from the Rasch analysis at the outset. Initial Rasch analysis of the remaining body structure categories showed a significant χ^2 ($p=0.006$), suggesting poor overall model fit. The PSI was 0.14 indicating poor internal consistency among the categories and inadequate ability to differentiate among the respondents along the trait being measured. Given this, no further analyses were performed for the ICF categories of body structures.

Activity and participation

The categories *d770 intimate relationships, d470 using transportation* and *d850 remunerative employment* had 29%, 34% and 69% missing values, respectively, and were excluded from the Rasch analysis at the outset. Initial analysis of the activity and participation component had a significant χ^2 ($p<0.001$) indicating a poor overall model fit (Table II, Analysis 4). The threshold map revealed extensively disordered thresholds. All categories (items) were rescored by collapsing response options (01234 to 01123). Although this improved the disordered thresholds there was still evidence of misfit among the items (Table II, Analysis 5).

On examination of the residual correlation matrix, *d445 hand and arm use* and *d440 fine hand use* were correlated at 0.52, indicating local dependency. These were sequentially deleted which resulted in a non-significant χ^2 ($p=0.02$) after a Bonferroni-adjustment to the α-value. The PSI was 0.77 (Table II, Analyses 6 & 7). Two categories *d450 walking* and *d455 moving around* were correlated at 0.37. Category *d450 walking* was then deleted, resulting in adequate overall fit (Table II, Analysis 8). No DIF with respect to age and sex were detected. However, significant uniform DIF on category *d920 recreation and leisure* for education was detected. Given the good overall
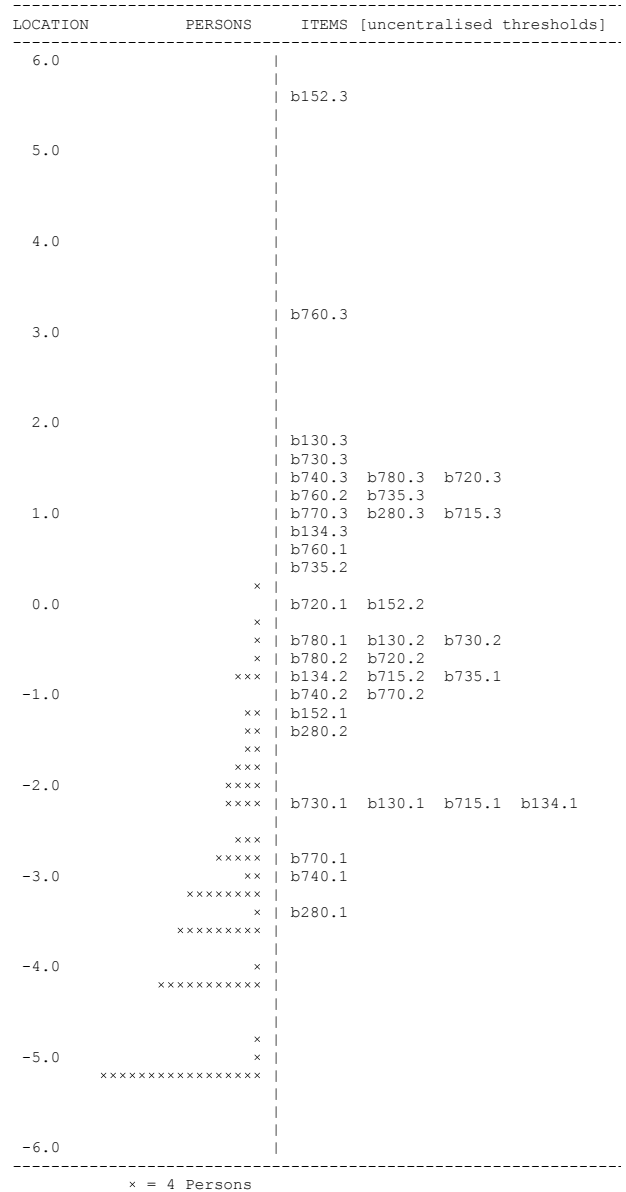
```
----------------------------------------------------------------
LOCATION       PERSONS      ITEMS [uncentralised thresholds]
----------------------------------------------------------------
  6.0                     |
                          |
                          | b152.3
  5.0                     |
                          |
                          |
                          |
  4.0                     |
                          |
                          |
                          | b760.3
  3.0                     |
                          |
                          |
                          |
  2.0                     |
                          | b130.3
                          | b730.3
                          | b740.3  b780.3  b720.3
                          | b760.2  b735.3
  1.0                     | b770.3  b280.3  b715.3
                          | b134.3
                          | b760.1
                          | b735.2
                   × |
  0.0                     | b720.1  b152.2
                  × |
                  × | b780.1  b130.2  b730.2
                  × | b780.2  b720.2
                ××× | b134.2  b715.2  b735.1
 -1.0                     | b740.2  b770.2
               ×× | b152.1
               ×× | b280.2
               ×× |
              ××× |
 -2.0        ×××× |
             ×××× | b730.1  b130.1  b715.1  b134.1
              ××× |
            ××××× | b770.1
 -3.0          ×× | b740.1
        ×××××××× |
               × | b280.1
        ××××××××× |
 -4.0          × |
        ×××××××××× |
                          |
               × |
 -5.0          × |
        ××××××××××××××××× |
                          |
                          |
 -6.0                     |
----------------------------------------------------------------
            × = 4 Persons
```

Fig. 1. Item map of the 12-item osteoarthritis (OA) core set body functions component for the participants with OA undergoing hip and knee joint replacement.

model fit and its clinical relevance, *d920 recreation and leisure* was retained and no further action was taken. Dimensionality testing revealed only 4 (4.17%) of the 96 *t*-tests with significant differences in scores from the two subsets of items, supporting unidimensionality of this component (Table II, Analysis 8).

The mean person location value was –4.42 (SD 1.94). The easiest categories were *d455 moving around, d640 doing housework,* and *d410 changing basic body position*. The most difficult categories were *d540 dressing, d660 assisting others,* and *d475 driving* (Fig. 2). Most categories were in the top half of the map (Fig. 2) and most participants had little difficulty endorsing the easy categories.

Osteoarthritis core set

The second stage in the analysis entailed integrating the retained ICF categories of the OA core set. Only the ICF categories of the body functions and activity and participation were suitable for combination as the body structures had very poor internal consistency making it inappropriate to be included as part of a potential tool for measuring function. Subtest analysis was performed by combining the retained items for the body function component to form one subtest, and the activity and participation component to form a second

```
--------------------------------------------------
LOCATION        PERSONS     ITEMS [uncentralised thresholds]
--------------------------------------------------
  6.0                       |
                            |
                            | d540.3
                            |
                            |
  5.0                       |
                            |
                            | d660.3
                            | d475.3
                            |
  4.0                       |
                            |
                            | d510.3
                            |
                            | d530.3
  3.0                       |
                            |
                            | d410.3
                            |
                            | d640.3
  2.0                       | d620.3  d510.2
                            |
                            | d910.3  d530.2
                            |
                            | d920.3  d415.3
  1.0                       |
                            |
                            |
                            | d455.3
                            | d540.2  d660.2
  0.0                       | d475.2
                            | d530.1  d430.3  d430.2
                            | d640.2  d620.2
                            | d415.2
                        × | d410.2  d510.1
 -1.0                       | d920.2  d910.2  d475.1
                        × |
                        × |
                        × | d660.1  d455.2
                            |
 -2.0                   × |
                        × |
                        × | d540.1
                        × | d910.1
                        × | d920.1
 -3.0                   × | d620.1
                        × |
                      ××× | d415.1  d430.1
                        × | d410.1
                       ×× |
 -4.0                       |
                       ×× |
                            | d640.1
                            |
                       ×× | d455.1
 -5.0                       |
                            |
                     ×××× |
                            |
                            |
 -6.0                       |
                            |
       ××××××××××××××× |
                            |
                            |
 -7.0                       |
--------------------------------------------------
         × = 7 Persons
```

Fig. 2. Item map of the 13-item osteoarthritis (OA) core set activity and participation component for the participants with OA undergoing hip and knee joint replacement.

subtest. Rasch analysis was conducted on these two subtests to assess their suitability to be combined to form a total score. Dimensionality testing revealed that only three cases out of 175 (1.71%) had statistically different scores on each subtest. This was below the cut-off point of 5% suggesting that the two separate reduced components could be integrated to form one measure (Table II, Analysis 9).

DISCUSSION

This is the first paper that investigates the potential of the ICF OA core set as a clinical measure of functioning in persons with OA undergoing hip and knee joint replacement using Rasch analysis. The study has two main findings. First, the ICF categories of the components activity and participation and body functions, but not body structures, could be integrated, reflecting a unidimensional construct, to represent a clinical measure of functioning in this population. However, removal of some categories and reduction of qualifier levels were necessary. Second, while the reduced ICF OA categories conform to a unidimensional construct, it was not well-targeted for this sample.

Unidimensionality

The ICF OA core set body functions, and activity and participation categories could be combined to form a clinical measure of functioning after omitting *b710 mobility of joint functions, d440 fine hand use, d445 hand and arm use*, and *d450 walking*. Category *b710* overlapped with other categories as any problems in joint range of motion lead to difficulties in other body functions, positions, and movements. For example, a limited knee range of motion or presence of contracture affect joint stability and gait pattern. On the other hand, *d440* and *d445* had little relevance as this study dealt more with lower extremity arthritis, and not many had upper extremity problems. The reason for the removal of *d450*, however, is less obvious. It does not mean that it is not important. From a clinical and conceptual point of view, it is; but from a measurement perspective, it is more discriminating than what the model predicts.

All participants responded similarly for all the categories across age and sex except for *d920 recreation and leisure* where some degree of uniform differential item functioning was evident for educational level. At equal levels of functioning, persons with tertiary education are more likely to endorse this category than those without tertiary education. This is probably a matter of resources and lifestyle preferences. Given good overall model fit, only marginal DIF and clinical relevance, this category was retained.

The categories from body structures could not be further analysed because these did not fulfil Rasch model expectations, as was also the case in another study (9). In contrast, another study found that most of the ICF categories (including body structures) comprising functioning reliably measured a single dimension (7). The main differences lie in the methodological approach and the study populations. One approach analysed the core set in its entirety at the outset by merging all the ICF categories of functioning (7). A similar method was used in other ICF studies
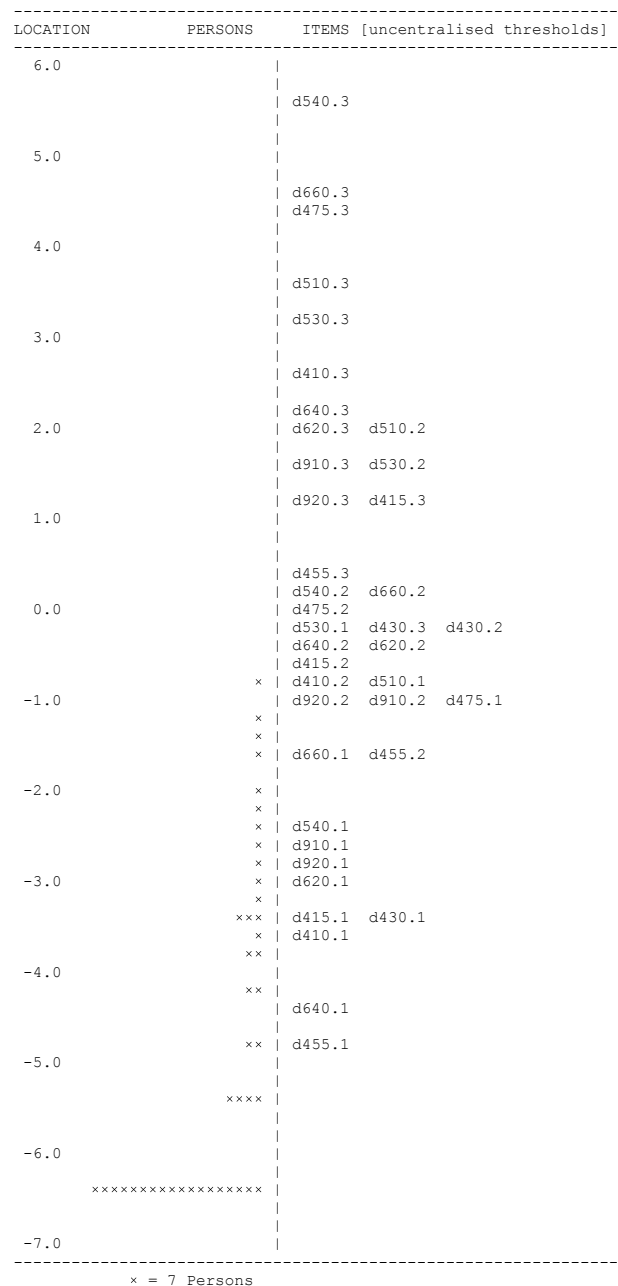
(8, 10). This current study presented another approach where Rasch analysis was performed in two stages. A two-staged analysis could potentially provide more information such as the relative contribution of components. The other dissimilarity is in the study population. Differences in health status, cultures, health care systems, and access would have an impact on results. While the other study examined data in Europeans and Asians with varying degrees of OA severity, this study included Australians with OA who were either waiting for, or within a year of, joint replacement surgery but with more in the latter. Thus, this sample, in general, probably had less severe problems in functioning.

Targeting

The uneven spread of the ICF categories across the full range of the participants' scores and the negative values for mean person location suggest poor item person targeting. A well-targeted measure would have items covering all the areas on the calibrated scale measuring the ability of all persons and would have a mean location of persons around the value of zero (16, 17). The item map of ICF categories for both body functions and activity and participation showed gaps as well as clustering of persons along the bottom margin. The gaps at the bottom of the map could imply lower reliability as the measure fails to differentiate persons along the latent trait. Furthermore, it could suggest the need to add new categories to discriminate between these persons of similar abilities. On the other hand, the clustering of persons at the bottom of the map signals the possibility of floor effect and also reflects presence of extreme scores. This is often observed after surgery and rehabilitation, where patients have improved functionally, leading to problems with detection of change. With floor and ceiling effects, future problems with responsiveness to change could be encountered. Bad targeting may also imply limited content relevance as items may not be adequate and relevant in describing the population in terms of the latent trait. Thus, mistargeting can cause a variety of measurement problems (20).

Suboptimal targeting in this case could be partly attributed to the nature of the study population. The sample consisted of more persons in the post-operative stage than pre-operative stage and might have already recovered with the joint problem addressed. Also, there were more participants seen in the private setting than in the public. The former may have had better access to health care services and did not have to wait for surgery, preventing further decline in health status and functioning.

Better targeting might be achieved when the measure is applied to this population pre-operatively and a few weeks post-operatively (sub-acute care stage). Alternatively, targeting could also possibly be improved by considering the addition of other relevant categories to the core set. Some studies have identified other pertinent ICF categories that include travelling, organized religion, weight maintenance, and looking after one's health (21, 22).

Clinical implications and limitations

Overall, using the current ICF taxonomy, the OA core set could represent a potential clinical measure assessing functioning

as a unidimensional construct in persons with OA undergoing arthroplasty. However, targeting was not optimal in fully evaluating the range of functioning and problems experienced in this particular sample. The ICF categories were more appropriate for use as a scale in persons with moderate to severe limitations in functioning.

The study has several limitations. First, only Australian data were included in the study. As the ICF is the framework for describing functioning and disability world-wide, this clinical measure should have international application. Further research, across other cultures, is required to confirm the findings of this study. Second, the sample might lack representativeness as it only covered approximately a third of the eligible population. There were more post-operative and private participants than pre-operative and public respondents. People who were feeling better and functioning well (e.g. able to walk, drive) were more inclined to participate in the study. Third, the removal of some ICF categories as a result of model misfit does not necessarily mean that these are irrelevant. These are important in describing functioning, goal-setting, and follow-up in rehabilitation. The clinical measure resulting from Rasch analysis does not replace the core set. It offers a means for the ICF to have more clinical and practical use in real world settings.

Conclusions

The results of the study suggest that the ICF OA core set could measure functioning as a unidimensional construct, with the removal of some categories and reduction in qualifier levels, in patients with osteoarthritis undergoing hip and knee joint replacement. The ICF categories are more suitable for use as a scale in persons with moderate to severe limitations in functioning. These findings demonstrate the potential of the ICF OA core set to be clinically useful in arthroplasty outcomes assessments. Further studies are needed to determine whether better targeting could be achieved in this population, particularly during the pre-operative stage and during the sub-acute care stage (rehabilitation) following surgery.

REFERENCES

1. Stucki G. International Classification of Functioning, Disability and Health (ICF): a promising framework and classification for Rehabilitation Medicine. Am J Phys Med Rehabil 2005; 84: 733–740.
2. Riddle D, Stratford P, Bowman D. Findings of extensive variation

in the types of outcome measures used in hip and knee replacement clinical trials. A systematic review. Arthritis Rheum 2008; 59: 876–883.

3. Alviar MJ, Olver J, Brand C, Hale T, Khan F. Do patient-reported outcome measures used in assessing outcomes in rehabilitation after hip and knee arthroplasty capture issues relevant to patients? Results of a systematic review and ICF linking process. J Rehabil Med 2011; 43: 374–381.

4. World Health Organization. International Classification of Functioning, Disability and Health: ICF. Geneva: WHO; 2001.

5. Cieza A, Ewert T, Üstün B, Chatterji S, Kostanjsek N, Stucki G. Development of ICF Core Sets for patients with chronic conditions. J Rehabil Med 2004; 36: 9–11.

6. Cieza A, Stucki G. The International Classification of Functioning, Disability and Health: its development process and content validity. Eur J Phys Rehabil Med 2008; 44: 303–313.

7. Cieza A, Hilfiker R, Chatterji S, Kostanjsek N, Ustün BT, Stucki G. The International Classification of Functioning, Disability, and Health could be used to measure functioning. J Clin Epidemiol 2009; 62: 899–911.

8. Prodinger B, Salzberger T, Stucki G, Stamm T, Cieza A. Measuring functioning in people with fibromyalgia based on the International Classification of Functioning, Disability and Health (ICF) – a psychometric analysis. Pain Practice 2012; 12: 255–265.

9. Roe C, Sveen U, Geyh S, Cieza A, Bautz-Holter E. Construct dimensionality and properties of the categories in the ICF core set for low back pain. J Rehabil 2009; 41: 429–437.

10. Cieza A, Hilfiker R, Boonen A, Van Der Heijde D, Braun J, Stucki G. Towards an ICF-based clinical measure of functioning in people with ankylosing spondylitis: a methodological exploration. Disabil Rehabil 2009; 31: 528–537.

11. Tesio L. Measuring behaviours and perceptions: rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003; 35: 105–115.

12. Dreinhofer K, Stucki G, Ewert T, Huber E, Ebenbichler G, Gutenbrunner C, et al. ICF core sets for osteoarthritis. J Rehabil Med 2004; 44: S75–S80.

13. Andrich D, Lyne A, Sheridan B, Luo G. RUMM 2030. Perth: RUMM Laboratory; 2010.

14. Rasch G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960.

15. Andrich D. Rasch models for measurement. Newbury Park, CA: Sage; 1988.

16. 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 (Arthritis Care & Research) 2007; 57: 1358–1362.

17. 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.

18. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. Br Med J 1995; 310: 170.

19. Smith E. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3: 205–231.

20. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. Int J Nurs Stud 2009; 46: 380–393.

21. Xie F, Thumboo J, Fong K-Y, Lo N-N, Yeo S-J, Yang K-Y, et al. Are they relevant? A critical evaluation of the international classification of functioning, disability and health core sets for osteoarthritis from the perspective of patients with knee osteoarthritis in Singapore. Ann Rheum Dis 2006; 65: 1067–1073.

22. Bossmann T, Kirchberger I, Glaessel A, Stucki G, Cieza A. Validation of the comprehensive ICF core set for osteoarthritis: the perspective of physical therapists. Physiotherapy 2011; 97: 3–16.