RESEARCH PAPER

Does the impact of osteoarthritis vary by age, gender and social deprivation? A community study using the International Classification of Functioning, Disability and Health

Beth Pollard[1], Diane Dixon[2], and Marie Johnston[1]

[1]Aberdeen Health Psychology Group, University of Aberdeen, Aberdeen, UK and [2]School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK

Abstract

Purpose: The aim of the study was to explore if the impact of osteoarthritis varies with respect to age, gender and social deprivation. Impact was defined as impairment, activity limitations and participation restriction (International Classification of Functioning, Disability and Health (ICF)). Investigating the functioning of the ICF model for subgroups is important both practically and theoretically. *Method*: The sample comprised a community sample of 763 people diagnosed with osteoarthritis. Uncontaminated measures of the ICF constructs were developed using discriminant content validity from a pool of 134 items, including the WOMAC and SF-36. Multigroup Structural Equation Modelling was used to explore if the same pathways exist for subgroups of gender, age and social deprivation. *Results*: Different significant paths were found for gender and social deprivation: impairment did not predict participation restriction for women and those most deprived, whereas these paths were significant for men and those less deprived. No difference in the paths was found for age. *Conclusions*: The impact of osteoarthritis appears to vary with respect to gender and social deprivation but not age. This suggests both that osteoarthritis *per se* does not adequately explain the health outcomes observed and that different clinical approaches may be appropriate for people of different gender and levels of deprivation.

➤ **Implications of Rehabilitation**

- The ICF model appears to vary with respect to gender and social deprivation for people with osteoarthritis.
- The ICF model did not appear to vary with respect to age for people with osteoarthritis.
- Different treatments and interventions for osteoarthritis may need to be targeted for specific gender and social deprivation groups.

Background

The dominant theoretical models of health outcomes or the consequence of disease have been the models developed by the World Health Organisation [1,2]. The most recent version, the International Classification of Functioning, Disability and Health (ICF) [1] is based on a biopsychosocial model that integrates medical and social models. The ICF model identifies three main distinct outcomes, Impairment (I), Activity limitations (A) and Participation Restrictions (P) and their respective opposites, Body Function and Structure, Activity and Participation.

The ICF model allows for all causal pathways between the constructs to be explored, with significant paths likely to vary for different conditions. While the relationships between I, A and P have been explored in some conditions, we are unaware of any studies that have explored if the relationships amongst the constructs are the same for different subgroups of a population. Investigating the functioning of the model for subgroups is important both practically and theoretically. For example, if we were to find that for people who are socially deprived, impairment directly impacts participation restriction, whereas for those who are less deprived there is no direct path, then it might be possible to target interventions more accurately and to gain greater understanding of the process of disablement.

In this study, we explore the ICF model by different subgroups in people with osteoarthritis (OA). There is evidence that for osteoarthritis there are strong relationships between impairment and activity limitation and between activity limitation and participation restriction with a weaker path between impairment

Address for correspondence: Beth Pollard, Aberdeen Health Psychology Group, School of Medicine and Dentistry, University of Aberdeen, 2nd Floor, Health Sciences Building, Foresterhill, Aberdeen AB25 2ZD, UK. Tel: 01224 438097. E-mail: bsp@st-andrews.ac.uk

and participation restriction [3–5]. However, subgroup comparisons have not been carried out.

We first removed two potential sources of measurement error. First, it has previously been shown that many existing measures do not map onto single ICF constructs (i.e. the items and/or measures are contaminated with other ICF constructs) [6,7], for example the item ''doing household tasks'' measures both activity limitations and participation restriction, making it impossible to examine relationships between A and P. For this study, we used the method of discriminant content validity (DCV) [6,8] to construct pure measures of impairment, activity limitation and participation restriction that were uncontaminated to enable us to accurately explore if the same model applies for subgroups of the population.

Second, it is also possible that different patterns of relationships could be found for different subgroups if there is measurement bias, i.e. items working in a different way for different subgroups. Such items exhibit differential item functioning (DIF) and DIF items have been found in health outcome measures [9–11]. Hence, we explored whether items had DIF to ensure that differences in models were not simply due to measurement bias. Thus, the aim of the study was to explore if the ICF model varies with respect to gender, age and social deprivation for people with osteoarthritis.

Participants and methods

Design

Statistical techniques were applied to an existing data set to explore if the ICF model varies with respect to gender, age and social deprivation. Uncontaminated, DIF-free measures of the ICF constructs were developed using DCV, classic and modern item analysis. Multigroup Structural Equation Modelling was used to explore if the same pathways exist for subgroups of gender, age and social deprivation.

Participants

The sample comprised a community of 763 people who had been diagnosed with osteoarthritis from 1359 people with hip and/or knee symptoms who completed a follow-up assessment (2002–2003) of health outcome measures as part of the Somerset and Avon Survey of Health Survey (SASH) [12,13]. Osteoarthritis was diagnosed by a clinician assessing X-rays using the Kellgren–Lawrence classification [14].

SASH is a large-scale survey of the population aged 35+. The age–sex stratified survey of 28 080 people registered with 40 general practices in Avon and Somerset yielded 2703 people reporting hip and/or knee symptoms at baseline (1994–1995). Written informed consent was obtained from all patients. Ethics approval was obtained from the South West Research Ethics Committee (MREC/01/6/51), and the study was conducted in accordance with the Helsinki Declaration.

Measures

ICF constructs

There was a pool of 134 potential items in the SASH database. These included two standard measures, the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [15], which is the most commonly used disease-specific measure of outcome used in OA. Also included was the Short Form-36 (SF-36) [16], which is the most commonly used generic measure of outcome used in OA [17]. However, these measures have been shown to contain items that map onto more than one

ICF construct [6,7] and have evidence of DIF for the groups investigated here [9,11,18–20].

Hence we constructed measures of impairment, activity limitation that were uncontaminated. Discriminant Content Validity Methods [6,8] were used to classify the pool of 134 items to I, A or P and item analysis was carried out in order to develop I, A and P measures. These items were then checked for DIF before examining relationships between variables. Where DIF items were identified, they were removed for the relevant analysis (see statistical analysis section).

Grouping factors

Gender, age and social deprivation (Townsend Index [21])

Median splits were used for age and social deprivation. The Townsend index was based on the four variables from the 1991 census: unemployment, non-car ownership, non-home ownership and household overcrowding.

Statistical analysis

Preliminary analysis: developing I, A and P measures

The following steps were carried out.

(i) DCV was used to classify the 134 self-report items to I, A and P or a combination thereof [6,8]. Six expert judges provided judgements of whether each item matched the theoretical definition of each ICF construct and provided a confidence rating for each judgement. Single sample *t*-tests were used to assess whether each item significantly tapped one unique construct or a mixture of constructs. Intraclass correlations (ICC) [22] were used to assess agreement between judges or inter-rater reliability applying a two-way mixed model with measures of consistency.

(ii) Item analysis was carried out on the resultant pool of unique I, A and P items from the DCV. A combination of classic psychometric methods and the modern methods of item response theory and DIF were used to develop the measures.

  (a) Items were removed if there was more than 10% missing data.

  (b) Items were removed that duplicated content.

  (c) Categories were collapsed if there was a less frequency of occurrence in this sample of less than 5%.

  (d) Cronbach's alpha was calculated for I, A and P with items removed if their inclusion decreased Cronbach's alpha.

  (e) Item response theory was carried out on the I, A and P measures to reduce items where there was more than 10 items in scale; in practice, this only applied to the activity limitation scale. Samejima's graded response model (GRM) [23] was fitted using MULTILOG [24]. Items with information <2.0 were removed.

  (f) Correlations between the resultant measures and other standard measures (SF-36 and WOMAC) were explored to evaluate validity. Cronbach's alpha was calculated to explore reliability.

  (g) DIF was carried out to evaluate if any of the items were biased for the subgroups under study, i.e. age group, gender and social deprivation group. The Swaminathan and Rogers (SR) [25] ordinal logistic regression method was used with a Bonferroni correction applied as this has been suggested to minimise Type 1 error due to the multiple testing [11,26].

Testing the model by grouping factor: multigroup structural equation modelling

Structural Equation Modelling (SEM) [27] is a statistical method that is mainly used to evaluate if a hypothesised theoretical model is plausible when compared to the observed data. SEM allows for the testing of both a measurement model and a structural model in a single model. Multigroup SEM allows for testing if the measurement and structural model are equivalent across groups. Multigroup SEM proceeds in hierarchical steps with increasingly restrictive models being tested by constraining sets of model parameters. Constraining models involves forcing the model's parameters to be the same for all groups: this evaluates if the same model holds between groups.

The main focus of this study was to use multigroup (SEM) to explore if structure equivalence exists by evaluating if there are similar path coefficients between the constructs for each subgroup. The I, A and P measures were used to explore the ICF model for each subgroup. The subgroups of gender, age, social deprivation were explored for the primary direction, i.e. I to A to P and from I to P.

As standard, one indicator factor loading was set to one, and initially correlations between the underlying latent factors were free to be estimated. As some items did not appear to be normally distributed, robust Maximum Likelihood estimation was used together with robust fit statistics and robust standard errors. The Satorra–Bentler chi-squared statistic [28] was calculated to assess model fit. As it has been shown that with large samples chi-square-based statistics are often highly significant even if there is good model fit [29], other fit indices were also explored. Model fit was assessed with emphasis on the robust comparative fit index (CFI), the normed fit index (NFI), the non-normed fit index (NNFI) and the robust Root Mean Squared Error of Approximation (RMSEA) with the 90% confidence interval. Fit indices greater than 0.90 have been considered satisfactory for model fit [30,31]. A RMSEA value of less or equal to 0.08 is generally accepted as an upper bound for acceptable fit [32]. If a small number of items (<5) prevented the model from converging, then these items were removed.

Nested models can be compared by calculating the differences in chi-square as this is also chi-square distributed with the degrees of freedom being the difference in degrees of freedom between the two models. However, if the non-normal Satorra–Bentler chi-squared statistic is used, then this does not apply [28] and a correction must be made. Satorra and Bentler have developed a method for computing a scaled difference chi-square statistic [33]. This was implemented using the Sbdiff.exe program developed by Prof. John Crawford, University of Aberdeen, (http://www.abdn. ac.uk/~psy086/dept/psychom.htm). Initially, a baseline 3 factor SEM model was fitted (including all the items i.e. without excluding any DIF items). For each subgrouping factor four nested models were fitted:

- **Model 1: 1 group baseline model**. The first model that was fitted was a baseline 3 factor (I, A, P) SEM model with the combined data, but omitting DIF items. Thus, Model 1 may vary from the initial baseline model if some items were identified from the DIF analysis and hence excluded for that particular subgroup analysis.
- **Model 2: 2 group baseline model**. The second model to be fitted was the baseline 3 factor two group SEM model tested simultaneously with no constraints. Good fit of this model would indicate configural equivalence, i.e. the number of factors and pattern of item loadings are similar across the groups
- **Model 3a: 2 group factor constrained model**. Thirdly, Model 2 was constrained by fixing the factor loadings to be equal across the two groups. If there is no significant difference between this model and Model 2, then the factor structure would appear to be similar between the two groups as this tests if the two models can be considered as having similar factor loadings. This would indicate measurement equivalence of the subgroups.
- **Model 3b: 2 group part factor constrained model**. If there was a significant difference between Model 3a and Model 2, then the Lagrange Multiplier (LM) test was used to identify the factor loading constraints that should be released to improve the measurement equivalence.
- **Model 4: 2 group factor and path constrained model**. Finally, Model 3 (a or b) was further constrained by fixing the path coefficients to be equal between the two groups to test the equality of the structural regression paths. A non-significant difference between this model and the Model 3 would indicate that the same model appears appropriate for each subgroup, i.e. the models can be considered as having similar path coefficients, and hence there is structural equivalence of the subgroups. If the models did appear to significantly differ, then the LM test was used to identify the paths that should be released to improve measurement equivalence, i.e. to identify the paths that were not equivalent.

Results

The participants demographic and WOMAC summary details are presented in Table 1.

Preliminary analysis: developing I, A and P measures

In brief, there was a pool of 134 potential items. DCV identified that 87 of these items were tapping only a single ICF construct (37 I items, 34 A items, 16 P items; ICC = 0.86). Of the 47 items that were not uniquely classified, 12 items were classified to both

Table 1. Participant characteristic table (percents or means and standard deviations and median split values for the grouping variables).

Gender (male)	43.3%
Age (years)	69.56 (9.84) Median = 70.88
Marital status (married)	72%
Ethnicity (white)	99.1%
Paid employment (yes)	33.7%
Social Class	
I	5.7%
II	33.8%
IIINM	19.0%
III M	23.4%
IV	14.7%
V	3.4%
Townsend quintiles (lower = most affluent)	Median = −1.47
20%	−2.76
40%	−1.89
60%	−0.80
80%	1.46
BMI	29.21 (5.30)
Affected joints	
Hip OA	487
Knee OA	612
Both Hip and Knee OA	336
Hip OA only	151
Knee OA only	276
No of affected joints	2.64 (1.01)
WOMAC physical	20.03 (14.51)
WOMAC pain for Hip OA	3.70 (4.16)
WOMAC pain for Knee OA	5.83 (4.34)
WOMAC stiffness Hip OA	1.70 (1.82)
WOMAC stiffness Knee OA	2.53 (1.92)

Table 2. Items in the final I, A and P measures.

Impairment items
  How much bodily pain in the past month
  Pain/Discomfort
  Pain during day over past year even with tablets/medication
  Pain during the night over past year even with tablets/medication
  Pain on walking over past year even with tablets/medication
  Have you had pain in past month when start to move or walk
  Do you have pain or discomfort during bed rest at night
  Do you experience morning stiffness or pain that decreases after rising
  Pain in last 48 hours at night while in bed
  How severe is the stiffness after sitting, lying or resting later in the day

Activity limitation items
  Does health limit you doing moderate activities
  Lifting or carrying groceries
  Climbing several flight of stairs
  Climbing one flight of stairs
  Maximum distance can you walk with pain
  Can you stand from sitting from a standard chair with no arms
  Can you walk on uneven ground
  Can you get in and out of a car
  Can you squat down
  Can you put on socks, stockings or socks by bending forward
  Can you pick up an object from the floor
  Are you able to get in and out of the bath
  Difficulty standing
  Difficulty bending to floor
  Walking on the flat
  Rising from bed

Participation restriction items
  Has health limited your social activities
  Do you have problems with usual activities
  Taking part in holidays or outings
  Going shopping

A and I, 2 to A and P with the remaining items unable to be classified to I, A or P. The item analysis, using classic test theory and item response theory, reduced the items to 10 I, 16 A and 4 P items. These items demonstrated good levels of reliability with Cronbach's alpha for I = 0.92, A = 0.96 and P = 0.75. There was also evidence of validity; the new measures were most highly correlated with the appropriate subscales from the established measures of SF-36 and WOMAC.

DIF analysis indicated that for Activity Limitation, 3 items (''limited lifting groceries'', ''problems getting in and out of the bath'' and ''maximum distance walk with pain'') were identified as having gender DIF and one item was identified as having age DIF (''difficulties putting on shoes, stockings, socks by bending forward''). Hence, for the analysis by gender and age, the respective DIF items were removed from the Activity Limitation measure. No DIF items were found for Impairment or Participation Restriction. The final measures are in Table 2.

Testing the model by factor: multigroup structural equation modelling
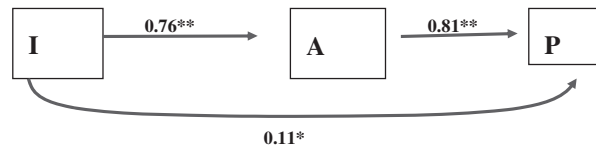
(a) Baseline overall model (with all measurement items):
The overall baseline model showed good fit. Overall significant paths were found between I, A and P with the strongest paths being between A and P (path coefficient = 0.83, $p < 0.01$) and I and A (path coefficient = 0.76, $p < 0.01$) with a much weaker but significant path, between I and P (path coefficient = 0.11, $p < 0.05$) (Figure 1a).
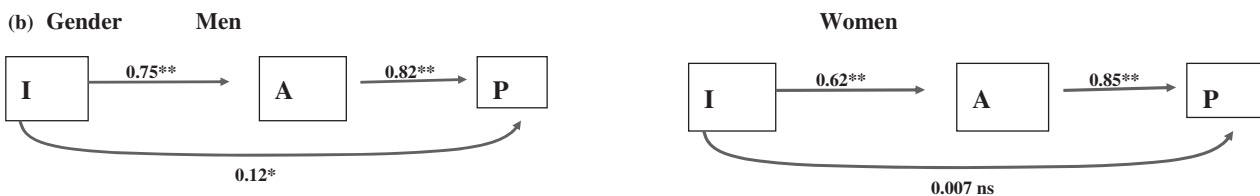
(b) Gender
For the exploration of the model by gender, the base 1 and 2 group models both showed good fit (Model 1 and Model 2) (Table 3; Figure 1b). The 2 group base model with factor loadings constrained to be equal (Model 3a) also showed good fit, but was
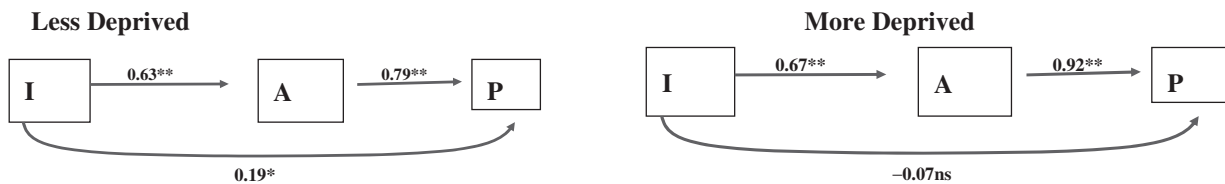

(a) Base one group model

(b) Gender   Men   **Women**

(c) Social deprivation

    Less Deprived   **More Deprived**

**p<0.01; *p<0.05, ns=non-significant

Figure 1. ICF structural paths between I, A and P. (a) Base one group model; (b) gender; (c) social deprivation.

Table 3 Summary table of the structural equation models for ''gender''.

Model		Chi-square	Robust Chi-Sq	df	CFI	NFI	NNFI	RMSEA	Chi square change (Satorra–Bentler)
Model 1	SEM base 1 group	3694.8	1515.7	227	0.979	0.975	0.976	0.092	
Model 2 (M2)	SEM base2 group	2339.0	1115.3	298	0.979	0.972	0.976	0.090	M2 and M3a = 47.54 (16)
									$p < 0.0005$
Model 3a (M3a)	Factor loadings constrained	2405.6	1166.8	314	0.978	0.970	0.976	0.089	M2 and M3b = 24.97 (15)
Model 3b (M3b)	With released factors if needed*	2377.2	1148.4	313	0.979	0.971	0.977	0.089	$p = 0.05$
Model 4 (M4)	Factor loadings and paths constrained	2433.4	1185.1	316	0.978	0.970	0.976	0.090	M3b and M4 = 179.62 (3) $p < 0.0005$

CFI: robust comparative fit index, NFI: robust normed fit index, NNFI: robust non-normed fit index, RMSEA: robust Root Mean Squared Error of Approximation.
*Constraint released – 1 item: ''difficulty standing''.

Table 4. Summary table of the structural equation models for ''social deprivation''.

Model		Chi-square	Robust Chi sq	df	CFI	NFI	NNFI	RMSEA	Chi square change (Satorra–Bentler)
Model 1	SEM base 1 group	4077.4	1712.3	296	0.983	0.979	0.981	0.085	
Model 2 (M2)	SEM base 2 group	2820.4	1346.3	412	0.983	0.975	0.981	0.083	M2 and M3a = 37.34 (19)
									$p = 0.007$
Model 3a (M3a)	Factor loadings constrained	2858.3	1383.7	431	0.982	0.975	0.981	0.082	M2 and M3b = 12.0 (17)
Model 3b (M3b)	With released factors if needed*	2840.7	1366.4	429	0.983	0.975	0.981	0.081	$p = 0.80$
Model 4 (M4)	Factor loadings and paths constrained	2894.8	1393.8	432	0.982	0.974	0.981	0.082	M3b and M4 = 27.41 (3) $p < 0.00005$

CFI: robust comparative fit index, NFI: robust normed fit index, NNFI: robust non-normed fit index, RMSEA: robust Root Mean Squared Error of Approximation.
*Constraints released for 2 items = ''holidays/outings'' and ''pain in bed at night''.

significantly worse than two group model without the constraints, indicating that not all the equality constraints were appropriate (Table 3). The LM test indicated that the constraint on the item ''difficulty standing'' should be released (as the factor loading for men appeared to be greater than for women (standardised path men = 0.83, women = 0.72). Model 3b was only just significantly different from Model 2, suggesting the measurement model (with one released constraint) was reasonable. The model with constrained structural paths added (Model 4) had worse fit and very significant chi-square difference from Model 3, and so was not structurally equivalent.

The path between I and P was significant for men (path coefficient = 0.12, $p < 0.05$), but not for women (path coefficient = 0.007, $p < 0.05$). The paths between A and P were similar for men and women (path coefficients = 0.82 and 0.85, respectively, $p < 0.01$). The LM test indicated that the path coefficients between I to A were significantly different for men and women (path coefficients = 0.75 (men) and 0.62 (women), $p = <0.01$) (Figure 1b).

(c) Social deprivation (Townsend: 2 groups with median split = −1.47)

For the exploration of the model by social deprivation group, the base 1 and 2 group models both showed good fit (Model 1 and Model 2) (Table 4; Figure 1c). The 2 group base model with factor loadings constrained to be equal (Model 3a) also showed good fit, but was significantly worse than the two group model without the constraints (Table 4). The LM test indicated that the constraint on the item ''holidays/outings'' and the item ''pain in bed at night'' should be released (with the factor loadings for those less deprived being greater than for those more deprived, 0.72 and 0.53 and 0.89 and 0.81, respectively). The model with these two released constraints (Model 3b) was now not significantly better than Model 2 and so indicated measurement equivalence. The model with constrained structural paths added (Model 4) had worse fit and very significant chi-square difference from Model 3b and so was not structurally equivalent.

The path between I and P was significant for those less deprived (path coefficient = 0.19 $p < 0.05$), but was not significant for those more deprived (path coefficient = −0.07, $p > 0.05$). When the LM test was examined, this path did not quite reach statistical significance for the difference ($p = 0.08$). The paths between I and A appeared similar for those less or more deprived (path coefficients = 0.63 and 0.67, respectively, $p < 0.01$). The paths between A and P were also both significant for those less or more deprived (path coefficients = 0.79 and 0.92, respectively, $p < 0.01$), but the LM test indicated that they were not significantly different from each other (Figure 1c).

(d) Age: 2 groups with median split = 70.88

For the exploration of the model by age group, the base 1 and 2 group models both showed good fit (Model 1 and Model 2, see Table 5). The 2 group base model with factor loadings constrained to be equal (Model 3) also showed good fit and was not significantly different to the two group model without the constraints (Model 2), hence this supports measurement equivalence. The model with the added structural path constraints (Model 4) was not significantly different from the model without the path constraints, so it appears that there was structural equivalence, i.e. the same model applies for both age groups.

Secondary analysis

As men and those less deprived had lower levels of impairment, activity limitation and participation restrictions, a *post-hoc* regression-based moderation analyses were carried out to explore if this would help explain the different models (based on median splits of I, A and P). However, these analyses did not offer evidence that the model significantly differed by level of I, A and P (not shown).

Discussion

Significantly different models were found for gender and social deprivation group but not for age. The paths between impairment

Table 5. Summary table of the structural equation models for ''Age''.

Model		Chi-square	Robust Chi-sq	df	CFI	NFI	NNFI	RMSEA	Chi square change (Satorra–Bentler)
Model 1	SEM base 1 group	3829.9	1578.8	272	0.983	0.979	0.981	0.085	
Model 2 (M2)	SEM base 2 group amend	2583.9	1218.3	372	0.983	0.976	0.981	0.083	M2 and M3a = 20.55 (18)
Model 3a (M3a)	Factor loadings constrained	2622.1	1243.4	390	0.983	0.975	0.982	0.081	$p = 0.30$
Model 4 (M4)	Factor loadings and paths constrained	2618.3	1246.3	392	0.983	0.975	0.982	0.081	M3a and M4 = 3.96 (3) $p = 0.27$

CFI: robust comparative fit index, NFI: robust normed fit index, NNFI: robust non-normed fit index, RMSEA: robust Root Mean Squared Error of Approximation.

and participation restriction were non-significant for women and those most deprived, but significant for men and those less deprived. Additionally, for men, the path between I and A was significantly stronger than for women. No significant difference between models was found for people of different ages although further splits could be explored in future work. *Post-hoc* analyses indicated that the effects of gender and deprivation were not simply due to different levels of disability (I, A, P) in the sub-groups, and hence unlikely to be due to the effects of co-morbidities. It appeared that the participation of men and those less deprived was more directly impacted by their impairment (without going through activity) than for women and those more deprived. Several explanations are possible. Participation restrictions that are not determined by activity limitations are more likely to be due to choices or decisions rather than being directly physically determined. Hence, it is possible that for men and those less deprived, the decision not to participate may be driven by different factors. On the one hand, groups may choose to cope with impairments using different coping strategies, e.g. by resting and avoiding exacerbation versus carrying on with tasks as usual. On the other hand, groups may differ in the nature of their tasks and responsibilities and how readily they can withdraw from them, e.g. it may be more difficult to withdraw from household tasks usually undertaken by women. This latter point might also explain why the path from I to A is weaker for women.

There is ample evidence that impairment may affect activity levels both directly and indirectly via individual goals, intentions and self-efficacy in performing activities [4,34–36]. Additionally, qualitative studies have identified a range of barriers to activities; Nicholson et al. [37] found that barriers perceived by stroke survivors mainly fell into three domains of the Theoretical Domains Framework, namely ''beliefs about capabilities'', ''environmental context and resources'' and ''social influence'' [37].

Thus, it might be that for women, these behavioural factors play a more important role relative to the biomedical impairment than they do for men. There are various reasons why women might show a different pattern of limitations given similar impairments to men's, including their role in the home and their response to the conditions [38].

The type of participation may differ not only by gender but also by social deprivation group, and therefore influence the strength of the I to P path; for example, those less deprived may be involved in work and leisure that is less reliant on activity limitation and hence their P may be directly impacted by I, for example our data indicated that those less deprived went to the cinema more than those more deprived. Also, as men and those less deprived drive more than women and those more deprived, their access to social and work activities may not be as dependent on their activity limitations. This may also explain the significant I to P paths and the very strong path between A and P for the more deprived group. Hence, the nature of I, A and P for gender and by social deprivation group needs further exploration. In particular,

it may be important to examine if there is a tipping point for degree of social deprivation or a qualitative difference in living style associated with deprivation that influences the impact of impairments on participation.

There were similarities between all the models with the paths from I to A, and from A to P being the strongest paths and a weaker path between I and P. This is consistent with previous studies of people with OA [3,5].

In practical terms, the results suggest that clinical interventions that reduce impairment might be expected to bring direct benefits in participation for men and those less deprived whereas other interventions, perhaps directed at reducing social burden or offering alternative methods of participation might be more helpful for women and those more deprived. Further, evidence of participation restrictions gives a less direct indication of problems in body structure or function for women and more deprived individuals.

The study has some limitations. We used cross-sectional data and cannot therefore draw conclusions about causal relationships. Even longitudinal data cannot test whether the relationships are truly causal and this will require experimental studies such as those conducted by Fisher and Johnston [39]. In this study we only explored I, A and P and not the contextual factors of the ICF. These contextual factors may have important moderating or mediating effects on the relationships between I, A and P. For example, our previous studies have found that psychological theories predict additional variance in A, beyond that explained by I, but have not investigated the impact of such variables on P [4].

The measures of I, A and P were derived and may need further validation and the number of items in the P measure was small and may not have fully covered the domain.

Thus, the impact of OA appears to vary with respect to gender and social deprivation but not age, in this sample. Thus, these groups may benefit from different treatments and interventions.

Declaration of interest

References

1. WHO. International classification of functioning, disability and health. Geneva: World Health Organisation; 2001.
2. WHO. International classification of impairments, disabilities and handicaps. Geneva: World Health Organisation; 1980.

3. Pollard B, Johnston M, Dieppe P. Exploring the relationships between International Classification of Functioning, Disability and Health (ICF) constructs of impairment, activity limitation and participation restriction in people with osteoarthritis prior to joint replacement. BMC Musculoskelet Disord 2011;12:97.

4. Dixon D, Johnston M, Pollard B. Using the ICF and psychological models of behavior to predict mobility limitations. Rehabil Psychol 2008;53:191–200.

5. Machado GPM, Gignac MAM, Badley EM. Participation restrictions among older adults with osteoarthritis: a mediated model of physical symptoms, activity limitations, and depression. Arthrit Rheumat 2008;59:129–35.

6. Pollard B, Johnston M, Dieppe P. What do osteoarthritis health outcome instruments measure? Impairment, activity limitation, or participation restriction? J Rheumatol 2006;33:757–63.

7. Weigl M, Cieza A, Harder M, et al. Linking osteoarthritis-specific health-status measures to the International Classification of Functioning, Disability, and Health (ICF). Osteoarthritis Cartilage 2003;11:519–23.

8. Johnston M, Pollard B. Consequences of disease: testing the WHO international classification of impairments, disabilities and handicaps (ICIDH) model. Soc Sci Med 2000;53:1261–73.

9. Pollard B, Johnston M, Dixon D. Exploring differential item functioning in the Western Ontario and Mcmaster universities osteoarthritis index (WOMAC). BMC Musculoskelet Disord 2012; 13:265.

10. Covic T, Pallant JF, Conaghan PG, Tennant A. A longitudinal evaluation of the Center for Epidemiologic Studies-Depression scale (CES-D) in a rheumatoid arthritis population using Rasch analysis. Health Qual Life Outcomes 2007;5:41.

11. Perkins AJ, Stump TE, Monahan PO, McHorney CA. Assessment of differential item functioning for demographic comparisons in the MOS SF-36 health survey. Qual Life Res 2006;15:331–48.

12. Eachus J, Williams M, Chan P, et al. Deprivation and cause specific morbidity: evidence from the Somerset and Avon survey of health. BMJ 1996;312:287–92.

13. Juni P, Dieppe P, Donovan J, et al. Population requirement for primary knee replacement surgery: a cross-sectional study. Rheumatology (Oxford) 2003;42:516–21.

14. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthrosis. Ann Rheum Dis 1957;16:494–502.

15. Bellamy N, Buchanan WW, Goldsmith CH, et al. 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 1988; 15:1833–40.

16. Ware JE, Sherbourne CD. The MOS 36-item short form health survey (SF-36). 1. Conceptual framework and item selection. Med Care 1992;30:473–83.

17. Bombardier C, Melfi CA, Paul J, et al. Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. Med Care 1995;33:AS131–44.

18. Bjorner JB, Kristensen TS. Multi-item scales for measuring global self-rated health: investigation of construct validity using structural equation models. Res Aging 1999;21:417–39.

19. Wolfe F, Hawley DJ, Goldenberg DL, et al. The assessment of functional impairment in fibromyalgia (FM): Rasch analyses of 5 functional scales and the development of the FM Health Assessment Questionnaire. J Rheumatol 2000;27:1989–99.

20. Davis AM, Badley EM, Beaton DE, et al. Rasch analysis of the Western Ontario McMaster (WOMAC) Osteoarthritis Index: results from community and arthroplasty samples. J Clin Epidemiol 2003; 56:1076–83.

21. Townsend P, Pkillimore P, Beattie A. Health and deprivation: inequality and the north. London: Routledge; 1988.

22. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1:30–46.

23. Samejima F. Estimation of a latent ability using a response pattern of graded scores. Psychometric Monograph 1969; Supplement No. 17. Richmond (VA): William Byrd Press.

24. Thissen D. MULTILOG: multiple, categorical item analysis and test scoring using item response theory. 6th ed. Chicago: Scientific Software; 1991.

25. Swaminathan H, Rogers HJ. Detecting differential item functioning using logistic regression procedures. J Educ Measurement 1990;26: 361–70.

26. Crane PK, Gibbons LE, Jolley L. Differential item functioning analysis with ordinal logistic regression techniques. DIF detect and DIF with par. Med Care 2006;44:S115–23.

27. Bryne BM. Structural equation modeling with EQS and EQS/Windows. Thousand Oaks (CA): Sage Publications; 1994.

28. Satorra A, Bentler PM. Corrections to test statistics and standard errors in covariance structure analysis. In: von Eye A, Clogg CC, eds. Latent variables analysis: applications for developmental research. Thousand Oaks (CA): Sage; 1994:399–419.

29. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990;107:238–46.

30. Bentler PM. On the fit of models to covariance and methodology to the bulletin. Psychol Bull 1992;112:400–4.

31. Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance-structures. Psychol Bull 1980;88: 588–606.

32. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Lond JS, eds. Testing structural equation models. Newbury Park (CA): Sage; 1993:136–62.

33. Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. Psychometrika 2001;66:507–14.

34. Dixon D, Johnston M, Elliott A, Hannaford P. Testing integrated behavioural and biomedical models of activity and activity limitations in a population-based sample. Disabil Rehabil 2012;34: 1157–66.

35. Quinn F, Johnston M, Johnston DW. Testing an integrated behavioural and biomedical model of disability in N-of-1 studies with chronic pain. Psychol Health 2013. [Epub ahead of print]. doi: 10.1080/08870446.2013.814773.

36. Quinn F, Johnston M, Dixon D, et al. Testing the integration of ICF and behavioral models of disability in orthopedic patients: replication and extension. Rehabil Psychol 2012;57:167–77.

37. Nicholson S, Donaghy M, Johnston M, et al. A qualitative theory guided analysis of stroke survivors' perceived barriers and facilitators to physical activity. Disabil Rehabil 2013; in press.

38. Mcgee HM, Hevey D, Johnston M, Pollard B. Does MI have a greater impact on women than men? Effects of gender role models and responsibilities. Eur Rev Appl Psychol 2000;50:333–40.

39. Fisher K, Johnston M. Experimental manipulation of perceived control and its effect on disability. Psychol Health 1996;11: 657–69.