



culty do you have standing?”, may measure pure activity limitation and none of the other domains. These possibilities were investigated using discriminant content validation methods in a recent study<sup>15</sup>. The study examined a range of instruments that are commonly used to assess health, including WOMAC<sup>14</sup> and the Medical Outcomes Study Short-form 36 (SF-36) health questionnaire<sup>16</sup>. A further measure, the low back pain Oswestry Disability Questionnaire (ODQ)<sup>17,18</sup>, was also investigated by the same authors using the same methodology (unpublished data). In these studies, expert judges classified individual items (questions) within the instruments as assessing 1 or more of the main ICF domains, i.e., each item was classified as measuring a pure domain (Impairment, Activity Limitation, or Participation Restriction) or a combination such as Activity Limitation and Participation Restriction. The agreement of judges on the classification was assessed by the intraclass correlation coefficient and found to be 0.95 for WOMAC, and 0.94 for both SF-36 and ODQ, suggesting very high agreement<sup>15</sup>.

It is important to use a reliable and valid instrument in trials, to accurately measure the effectiveness of an intervention. We aimed to investigate whether measures that assess the individual domains of the ICF model are more sensitive than the conventional measures that assess a mixture of the different domains. We aimed to test the hypothesis that different types of interventions might have different effects on impairment (I), activity limitation (A), and participation restriction (P). If true, then using global measures that include items from different domains may be inappropriate, and inadequate in inferring the true efficacy of different interventions.

## MATERIALS AND METHODS

This is a secondary analysis of data from 3 previously published randomized controlled trials (RCT) conducted in the UK. Two were done on patients with knee OA and 1 on patients with sciatica. The 3 studies were chosen on the basis of convenience and availability of raw data. In each trial, 1 or 2 of the outcome measures (WOMAC function, SF-36 physical function, and ODQ) were used for the reassessment of the effect of the different interventions. The 3 measures included mostly pure A and P items or a combination of A/P, hence the investigation was confined to these and do not include I. The chosen outcome measures had previously been judged on whether they fit with the ICF structure<sup>15</sup>. For example, the 17 items of WOMAC function were reclassified as a combination of 12 pure Activity Limitation (A) items, and 5 questions judged to assess a mixture of Activity Limitation and Participation Restriction (A/P; Table 1). We compared the originally reported outcomes using total WOMAC function (17 items) with the newly defined subscales of 12 pure items on A and 5 mixture items (A/P). Summation over the appropriate items was used. Details of the classification of items for the outcome measures were given in Table 2 of Pollard, *et al*<sup>15</sup>.

The 3 trials examined the efficacy of a range of interventions including a pharmacological package, specific medical procedures and injections, and physiotherapy.

*The KIVIS Study (knee intraarticular therapy trial)*. This was a randomized, controlled, single-blind, dual-center trial of 150 patients, aged between 40 and 90 years, with knee OA. Patients were randomized to 2 par-

Table 1. Classification of SF-36 physical function, WOMAC function, and Oswestry Disability Questionnaire (ODQ) items into subdomains.

Short Form (SF-36) Questionnaire	
Physical Function Domain	
Activity and Participation (“A/P”)	Activity only (“A”)
Vigorous activities	Climbing several flights of stairs
Moderate activities	Climbing one flight of stairs
Lifting and carrying	Bending, kneeling or stooping
	Walking more than a mile
	Walking half a mile
	Walking 100 yards
	Bathing and dressing self
WOMAC Function	
Activity and participation (“A/P”)	Activity only (“A”)
Getting in/out of car	Descending stairs
Going shopping	Ascending stairs
Getting in/out of bath	Rising from sitting
Heavy domestic duties	Standing
Light domestic duties	Bending
	Walking
	Putting on socks/shoes
	Taking off socks/shoes
	Sitting
	Rising from bed
	Lying in bed
	Getting on/off toilet
Oswestry Disability Questionnaire	
Participation only (“P”)	Activity only (“A”)
Social life	Personal care
Traveling	Lifting
	Walking
	Sitting
	Standing

SF-36: Medical Outcomes Study Short Form 36 questionnaire; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Measure.

allel groups. Seventy-one received tidal irrigation (TI) using a 3.2 mm arthroscope under local anesthesia, and 79 received an intraarticular injection of corticosteroid (SI)<sup>19</sup>. The original study used the visual analog scale (VAS) of WOMAC function, in which responses are measured on a 0–10 scale, and expressed as percentages. The final score was obtained using the standard procedure, adding up all items. A single score of WOMAC function was calculated at 2, 4, 12, and 26 weeks and compared with the baseline score for each group. We further used 2 subscales to assess changes in A only and changes in A/P separately using similar procedures; t-test was used for all comparisons.

*The Wessex Epidural Steroids Trial (WEST)*. In this trial, 228 patients aged 18–70 years with a clinical diagnosis of unilateral sciatica (1–18 months duration) were randomized to either placebo (injections of 2 ml of normal saline into the interspinous ligament) or a lumbar epidural injection of 80 mg triamcinolone acetonide and 10 ml of 0.25% bupivacaine corticosteroid at Weeks 0, 3, and 6. The patients were assessed at 0, 3, 6, 12, 26, and 52 weeks. The primary outcome measure was the ODQ. Secondary outcomes included the SF-36 physical function subscale and other walking and climbing time measures<sup>20</sup>. We reexamined changes in A only and in A/P based on the SF-36 physical function subscale and changes in A and P only based on the ODQ subscales<sup>15</sup> for the 2 groups. For the ODQ, scores for A only and P only were calculated for each domain using the standard formula used for the full scale, and changing the number of items as appropriate for each subscale; t-test was used for all comparisons, as no difference in results was reported where other regression methods were attempted.

*The TOPIK trial (Treatment Options for Pain in the Knee).* This was a pragmatic multicenter RCT undertaken in 15 general practices in North Staffordshire, UK<sup>21</sup>. Participants were 325 adults aged ≥ 55 years (mean 68 years) consulting with knee pain; 297 (91%) reached a 6-month followup. Interventions were enhanced pharmacy review (pharmacological management according to an algorithm); community physiotherapy (advice about activity and pacing and an individualized exercise program); and control (advice leaflet reinforced by telephone call). The main outcome measure was change in WOMAC function at 3, 6, and 12 months (Likert version). Similar procedures were adopted in this investigation to examine changes in A only and A/P only items.

*Analysis.* Analysis was by intention to treat (ITT) as adopted in the original studies. All outcome measures were treated as last observation carried forward so that data were available for every subject at each timepoint, wherever that was possible. Estimates of the treatment effects with 95% CI, based on t-test or regression methods, were used as appropriate. The same procedures were used for full scale and subscales at all times. Original findings based on the global outcome measures were reproduced, and in addition the newly defined subscales such as A only, P only, and A/P were then used as outcome measures with the same statistical techniques. For all comparisons the effect size was calculated as the difference between means/combined SD.

Results for the newly defined subscales of A only, P only, and A/P are presented with the original results based on full scales in Table 2, with subsections (1), (2), (3), and (4) for the different scales examined. The results in Table 2 (2) are based on SF-36 physical function scores using standard procedures<sup>16</sup> for the full 10-item scale. We adjusted the number of items as appropriate for the A and A/P subscales. We used the t-test to compare change in score in the 2 groups.

## RESULTS

*KIVIS [Table 2 (1)].* The original study reported reduction in

pain and improvement in function in both groups at Weeks 2 and 4, but no significant difference in benefit between the 2 interventions. At Weeks 12 and 26, the benefit of SI decreased while that of TI was maintained, and this difference was significant at the 5% level. When we examined A (only items based on WOMAC function scale in comparison with A/P items), the improvement in A/P from TI was more pronounced than that on A only. The effect size for A only was 0.01, 0.04, 0.32, and 0.38 at Weeks 2, 4, 12, and 26, respectively, and the corresponding effect sizes for the A/P items were 0.08, 0.16, 0.38, and 0.51. This suggests that TI might have affected participation restriction more than activity limitation.

*The WEST Study [Table 2 (2 and 3)].* We examined changes in the SF-36 physical function score for the placebo and the lumbar epidural corticosteroid (ESI) treatment groups at 3, 6, 12, 26, and 52 weeks. All 10 items of the domain were used as a global measure. A only (7 items) and A/P (3 items) were also used separately. The SF-36 physical function as a global score did not detect a significant response to treatment compared to placebo, as reported in the original study<sup>20</sup>. The change in SF-36 physical function scores, however, suggested an improvement in physical function for the 2 groups at all assessments. This improvement was greater for the ESI group than for placebo in the first 2 assessments but smaller after that (Weeks 12, 26, and 54). When the subscales A only and A/P were examined separately, the place-

Table 2 (1). KIVIS study. Changes in WOMAC function from baseline to 2, 4, 12, and 26 weeks, for all items, A only, and A/P items for 2 treatment groups: tidal irrigation (TI) and corticosteroid injection (CSI). Comparisons are based on an unadjusted regression model.

No. of Subjects	Baseline Values		Weeks							
	TI	CSI	2		4		12		26	
			TI	CSI	TI	CSI	TI	CSI	TI	CSI
	71	79	70	77	69	77	68	73	65	71
	Total Score (SD)		Decrease from Baseline (SD)							
WOMAC function										
All items (17)	853	831	264	259	284	262	255	144	219	69
Scores (SD)	(312.6)	(340.7)	(295)	(254)	(301)	(292)	(333)	(294)	(356)	(328)
Difference in mean (95% CI)			-5.6 (-94.1, 83.0)		-22.5 (112, 73)		-111 (-212.5, -10.1)		-15 (-260.8, -40.2)	
Effect size			0.02		0.08		0.33		0.43	
p			0.90		0.64		0.03		0.01	
Scores (SD)	582	573	187	188	195	186	174	100	149	53
A only items (12)	(220)	(243)	(217)	(184)	(234)	(214)	(243)	(213)	(256)	(238)
Difference in mean (95% CI)			0.96 (-63.8, 65.7)		-9.0 (-81.4, 63.4)		-74 (-147.7, -0.71)		-96.1(-175.8, -16.4)	
Effect size			0.01		0.04		0.32		0.38	
p			0.98		0.81		0.05		0.02	
Score (SD)	271	259	78	71	89	76	81	44	70	15
A/P items (7)	(104)	(105)	(57)	(82)	(69)	(87)	(102)	(89)	(111)	(97)
Difference in mean (95% CI)			-6.7 (-34.2, 20.9)		-13.8 (-41.5, 14.0)		-37.2 (-68, -6.4)		-54.5 (-88.2, -20.9)	
Effect size			0.08		0.16		0.387		0.51	
p			0.63		0.33		0.02		0.002	

WOMAC: Western Ontario and McMaster Universities Osteoarthritis Measure based on a visual analog scale. A: activity limitation; P: participation restriction.

Table 2 (2). The WEST study. Changes in SF-36 physical function scores at 3, 6, 12, 26, and 52 weeks, for all items, A only, and A/P items, for 2 treatment groups. Comparisons are based on t-tests.

	Absolute Score (Total)		Change from Baseline		Mean Difference (95% CI)
	ESI	Placebo	ESI	Placebo	
All (10 items)	120	1087	120	108	
Baseline	18.09	17.75			
3 months	20.36	19.77	2.26	2.02	0.24 (−0.79, 1.28)
Effect size					0.06
p					0.64
6 months	21.2	20.5	3.11	2.75	0.35 (−0.83, 1.54)
Effect size					0.08
p					0.56
12 months	20.88	21.03	2.78	3.28	−0.50 (−1.88, 0.89)
Effect size					−0.09
p					0.48
26 months	20.79	20.83	2.70	3.09	−0.39 (−1.85, 1.07)
Effect size					−0.07
p					0.74
52 months	23.23	23.13	5.13	5.38	−0.25 (−1.75, 1.25)
Effect size					−0.04
p					0.74
A only (7 items)					
At baseline	13.63	13.27			
3 months	15.13	14.87	1.50	1.60	−0.10 (−0.70, 0.91)
Effect size					−0.03
p					0.80
6 months	15.77	15.44	2.14	2.17	−0.03 (−0.86, 0.93)
Effect size					−0.01
p					0.94
12 months	15.50	15.71	1.87	2.45	−0.58 (−0.44, 1.59)
Effect size					−0.15
p					0.26
26 months	15.38	15.48	1.75	2.22	−0.46 (−0.62, 1.54)
Effect size					−0.11
p					0.40
52 months	16.96	17.07	3.33	3.81	−0.48 (−0.62, 1.58)
Effect size					−0.11
p					0.39
A/P (3 items)					
At baseline	4.45	4.49			
3 months	3.27	3.23	−1.18	−1.25	0.07 (−0.23, 0.38)
Effect size					0.06
p					0.63
6 months	5.43	5.06	0.98	0.58	0.40 (−0.02, 0.82)
Effect size					0.25
p					0.06
12 months	5.37	5.31	0.92	0.83	0.09 (−0.41, 0.59)
Effect size					0.05
p					0.72
26 months	5.41	5.35	0.95	0.87	0.09 (−0.40, 0.58)
Effect size					0.05
p					0.72
52 months	6.27	6.06	1.81	1.57	0.24 (−0.29, 0.78)
Effect size					0.12
p					0.37

WEST: Wessex Epidural Steroids Trial; SF-36: Medical Outcomes Study Short-form 36 questionnaire; ESI: epidural steroids injection; A: activity limitation; P: participation restriction.

bo group did better in A only items at all assessments (−0.03, −0.01, −0.15, −0.11, and −0.11), while the ESI group did better than placebo for A/P items in all assessments including

the first, where the 2 groups showed some decline (0.06, 0.25, 0.05, 0.05, and 0.12). The magnitude of decline in the first assessment was smaller for the ESI group. These results

Table 2 (3). The WEST study. Changes in Oswestry Disability Scores from baseline to followup, on all items, A only, and P only items, for the treatment (ESI) and the placebo (PL) groups. P values were based on ANOVA adjusting for baseline scores.

	Baseline Score (SD)		Change in Score Between Baseline and Weeks 3, 6, 12, 26, and 52 (SD)									
	ESI	PL	3		6		12		26		52	
	ESI	PL	ESI	PL	ESI	PL	ESI	PL	ESI	PL	ESI	PL
All items, score (SD)	44 (15)	45 (18)	10 (15)	7 (16)	13 (17)	10 (18)	12 (19)	12 (21)	12 (20)	11 (22)	21 (22)	20 (24)
Mean difference			3.7		2.5		0.25		0.74		1.57	
(95% CI)			(-0.25, 7.6)		(-2.1, 7.2)		(-4.9, 5.4)		(-4.6, 6.1)		(-4.5, 7.6)	
Effect size			0.23		0.14		0.01		0.04		0.07	
p			0.052		0.60		0.97		0.58		0.33	
A only	41 (15)	42 (18)	7.9 (13)	5.2 (16)	10 (16)	8 (18)	9 (17)	9 (20)	9 (18)	9 (21)	17 (21)	17 (23)
Mean difference			2.6		2.0		-0.17		0.48		0.43	
(95% CI)			(-1.2, 6.5)		(-2.4, 6.5)		(-4.9, 4.6)		(-4.6, 5.6)		(-5.2, 6.1)	
Effect size			0.18		0.12		-0.01		0.02		0.02	
p			0.05		0.32		0.81		0.60		0.37	
P only items	47 (21)	49 (26)	13 (20)	8 (21)	16 (23)	12 (23)	14 (25)	15 (27)	14 (25)	13 (26)	25 (27)	23 (32)
Mean difference			4.4		3.7		-0.21		0.9		1.5	
(95% CI)			(-0.95, 9.8)		(-2.4, 9.8)		(-7.0, 6.6)		(-5.9, 7.6)		(-6.1, 9.2)	
Effect size			0.21		0.16		-0.01		0.03		0.05	
p			0.01		0.05		0.38		0.28		0.12	

WEST: Wessex Epidural Steroids Trial; A: activity limitation; P: participation restriction; ESI: epidural steroids injection.

suggest that while the global scale showed no change and some inconsistency, the subscales have detected benefit for the treatment on A/P. While the effect size for A only was negative in the 5 assessments, the corresponding effect sizes for A/P items were all positive, pointing to the benefit of the treatment over placebo. The original study showed that when ODQ was used as a global outcome measure, there was a transient benefit for ESI over placebo at 3 weeks but no difference after that. When the measure was reexamined as A only and P only items, the effect size for P items was found to be slightly larger than that for A in the first 2 assessments, suggesting a possible differential effect of the active intervention, in agreement with the finding in the SF-36 physical function scores.

*The TOPIK Trial [Table 2 (4)].* The study addresses the hypothesis that different types of intervention have a different effect on A versus P, as there are 2 quite different interventions: enhanced pharmacy and physiotherapy. The original study<sup>21</sup> showed a significant improvement in WOMAC pain scores for both the pharmacy and the physiotherapy groups compared with control at the 3-month assessment, but no significant difference for any of the interventions at the subsequent 6-month and 12-month assessments. For WOMAC function, the original study showed a significant improvement in the scores for the physiotherapy group at 3 months, and a slight but not significant improvement for the pharmacy group.

When we reexamined the outcome measure as subscales A and A/P, at 3 months there was a positive effect from pharmacy on both A and A/P, with a larger effect size for A than for A/P, while the effect of the physiotherapy was larger for A/P than for A. The effect of the pharmacy, however, decreased after 3 months, and the control group did better at

6 and 12 months. At both assessments the difference between the 2 groups was very small and not significant, and the difference on A/P was negligible. On the other hand, the effect of physiotherapy was higher on A/P at 6 months, and that was maintained at 12 months, suggesting a consistent improvement on P (although not significant at the 5% level) at both occasions.

## DISCUSSION

We examined the differential effect of a range of interventions on different domains of health as identified by the ICF model<sup>1</sup>, using 3 outcome measures: SF-36 physical function, WOMAC function, and ODQ. Each item of these instruments was classified as assessing 1 or more domains, i.e., each item was classified as measuring a pure domain (impairment, activity limitation, or participation restriction), or mixed, such as activity limitation and participation restriction<sup>15</sup>. The study supports the hypothesis that different interventions may differentially affect different health domains. Although OA and back pain with sciatica were the focus of this study, the principle of examining different domains of health probably relates to many common musculoskeletal conditions.

The original KIVIS study showed positive effects for TI over corticosteroid injection, with TI significantly improving pain and function. When we examined the subscales of WOMAC function, the effect size for A/P items was larger than that for A only items, suggesting the possible effect of improvement in pain on participation. A differential effect was also noted in the TOPIK study, in which physiotherapy had a larger effect on participation restriction, while enhanced pharmacy seemed to improve activity more. Similarly in the WEST study, using A only items from the

Table 2 (4). TOPIK trial. Changes in WOMAC function score, all items, A items only, and A/P items, for enhanced pharmacy and physiotherapy treatment groups, and a control group. Comparisons were based on ANOVA, adjusting for baseline scores. Estimates were rounded to the nearest 1 decimal place.

WOMAC Function	Control (n = 90)	Enhanced Pharmacy Review Scores (SD) (n = 97)	Community Physiotherapy Mean Difference (95% CI)	Control Scores (SD) (n = 95)	Enhanced Pharmacy Review Mean Difference (95% CI)
3 months (All)					
All items (17)					
Absolute score; mean SD	30.2 (12.8)	25.7 (13.4)		24.3 (15.2)	
Crude change score	0.8 (8.5)	2.61 (9.8)	1.80 (-0.8, 4.5)	4.8 (10.8)	3.99 (1.2, 6.8)
Effect size			0.20		0.40
p			0.10		0.01
Adjusted change score			2.12 (-0.5, 4.8)		3.65 (1.0, 6.3)
A only (12 items)					
Absolute score	20.7 (9.3)	17.23 (9.6)		16.44 (10.8)	
Crude change score	0.59 (6.18)	1.95 (7.14)	1.36 (-0.57, 3.30)	3.13 (7.85)	2.54 (0.48, 4.60)
Effect size			0.20		0.34
p			0.17		0.02
Adjusted change score			1.66 (-0.24, 3.58)		2.39 (0.46, 4.33)
A/P items (5 items)					
Absolute score	9.43 (3.88)	8.44 (4.20)		7.83 (4.59)	
Crude change score	0.22 (3.04)	0.69 (3.04)	0.47 (-0.41, 1.36)	1.69 (3.33)	1.47 (0.55, 2.39)
Effect size			0.16		0.45
p			0.29		0.002
Adjusted change score			0.50 (-0.37, 1.37)		1.26 (0.40, 2.13)
6 months (All)					
All items (17)					
Absolute score	28.2 (13.2)	26.8 (13.4)		25.5 (16.3)	
Crude change score	2.74 (10.5)	1.52 (11.4)	-1.23 (-4.4, 1.9)	3.34 (12.2)	0.59 (-2.7, 3.9)
Effect size			-0.11		0.05
p			0.5		0.7
Adjusted change score			-0.96 (-4.0, 2.1)		0.66 (-2.5, 3.8)
A only (12 items)					
Absolute score	19.24 (9.23)	18.24 (9.28)		17.57 (11.58)	
Crude change score	1.92 (7.43)	0.99 (7.86)	-0.93 (-3.12, 1.27)	1.88 (9.00)	-0.04 (-2.42, 2.33)
Effect size			-0.12		-0.01
p			0.40		0.97
Adjusted change score			-0.69 (-2.78, 1.39)		0.11 (-2.15, 2.37)
A/P items (5 items)					
Absolute score	8.90 (4.18)	8.63 (4.36)		7.98 (5.0)	
Crude change score	0.82 (3.58)	0.63 (3.90)	-0.19 (-1.27, 0.90)	1.47 (3.64)	0.65 (-0.39, 1.69)
Effect size			-0.05		0.18
p			0.73		0.22
Adjusted change score			-0.17 (-1.21, 0.87)		0.59 (-0.42, 1.60)
12 months (All)					
All items (17)					
Absolute score	29.0 (14.1)	27.0 (14.4)		25.0 (15.3)	
Crude change score	1.65 (12.3)	1.15 (11.7)	-0.50 (-4.0, 2.8)	4.0 (13.2)	2.5 (-1.2, 6.1)
Effect size			-0.05		0.19
p			0.80		0.20
Adjusted change score			-0.39 (-3.8, 3.0)		2.41 (-1.1, 5.9)
A only (12 items)					
Absolute score	19.86 (10.1)	18.48 (10.32)		16.55 (10.78)	
Crude change score	1.13 (8.75)	0.67 (8.52)	-0.44 (-2.97, 2.10)	2.77 (9.58)	1.64 (-1.05, 4.33)
Effect size			-0.05		0.18
p			0.73		0.23
Adjusted change score			-0.28 (-2.70, 2.15)		1.84 (-0.62, 4.3)
A/P items (5 items)					
Absolute score	9.09 (4.57)	8.68 (4.53)		8.10 (4.67)	
Crude change score	0.51 (4.09)	0.47 (3.63)	-0.04 (-1.17, 1.09)	1.29 (4.12)	0.78 (-0.43, 1.98)
Effect size			-0.01		0.19
p			0.94		0.21
Adjusted change score			-0.05 (-1.15, 1.06)		0.66 (-0.46, 1.78)

TOPIK: Treatment Options for Pain in the knee; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Measure; A: activity limitation; P: participation restriction.

SF-36 physical function subscale detected benefit from the epidural steroid injection (ESI) over placebo, which was not apparent when the full physical function scale was used. Although not significant at the 5% level, these results are of interest since the positive effect size for A/P, as opposed to negative effect size for A only items, provides some evidence for a differential effect. The ESI appears to have improved participation but not activity. Thus, combining subscales that measure different domains may result in inconsistencies that disappear once they are treated separately. In such situations, pooling items could result in cancellation of opposite sign estimates (negative and positive) and eventually any effect would disappear or at least get diluted in the final total figure. Thus true treatment effects might be masked by using global outcome measures. The original WEST study showed a significant improvement in pain in the first assessment, and a slightly better score was also reported for the (ESI) group after that, supporting the possibility of better ability to participate. The possible relationships between the interventions investigated and the ICF domains are described in Figure 1.

Interest in outcome measures has occupied a central position in many fields of clinical and methodological research, and various issues have been addressed. The contents of measures and their suitability to assess different health conditions has been questioned by Hunt and McKenna<sup>22</sup>, and their responsiveness to change was highlighted by Ware, *et al*<sup>23</sup>. Methodological problems that arise in attempting to use simple algebraic summation to aggregate items that may measure different domains of health has been explored by many researchers<sup>24,25,26</sup>. Methods such as factor analysis and principal components are widely used to identify underlying structures of many health outcome measures. An example is the General Health Questionnaire<sup>27,28</sup>. The complexity of these methods, however, and occasional inappropriate use have resulted in much confusion and inconsistency in the inferences drawn<sup>29</sup>. Latent class models that solve many of the methodological problems of these classical methods were recently developed<sup>30,31</sup>.

In addition to methodological issues, however, these statistical methods do not readily match measures to the theo-

retical domains of the ICF, but attempt to empirically identify dimensionality, and these methods cannot map measurement items to the theoretical domains such as those identified in the ICF. An alternative is the linkage method suggested by Cieza, *et al*. For example, they have indicated methods of linking OA measures to the ICF categories<sup>2,32,33</sup>. However, they have not attempted to separate A and P, finding this too difficult using the linkage method. In our study, the discriminant content validation methods involving experts' judgment rather than mathematical models, used to identify items as uniquely measuring I, A, or P, or any combination of these (Pollard, *et al*<sup>15</sup>), were numerically investigated. Our results show that it is indeed possible to find significant differences in attributing items to A or P<sup>15</sup>, and the results suggest that separating A and P may be important in investigating the effectiveness of different interventions. The view that A and P are combined, suggested by the ICF categories, was often challenged by researchers. Instead of linking items to the ICF categories, the hypothesis that the 2 concepts are distinct was tested. In a study assessing physical function, on community adults, for example, 3 underlying structures were identified, and the authors concluded that combining A and P might not be appropriate<sup>34</sup>.

Our study addressed the hypothesis that different types of intervention used to treat OA or sciatica might have differential effects on activity limitation and restricted participation. Drugs were found to affect activity limitation more than participation restriction, and exercise and injection have more effect on participation restriction than on activity limitation. Different methods of classification would have obtained different conclusions. For example, based on the ICF categories, A and P would have been treated as 1 component. Based on Badley<sup>35</sup>, what was defined as acts such as walking and tasks such as dressing would have been considered different, and different conclusions would have been reached. The results from our study, although not conclusive, may help answer interesting questions raised by Johnston and Pollard<sup>36</sup> regarding the ability of existing measures to separate individual structures of health, and may provide evidence for the effectiveness of interventions that was not detected by global outcome measures. Such

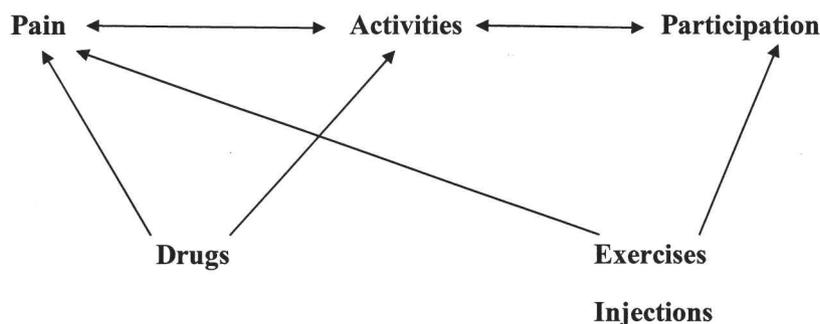


Figure 1. Possible effects of different interventions on individual structures of the International Classification of Functioning, Disability and Health model.

evidence may be of value in advising treatment and rehabilitation programs. Clinical implications, however, are yet to be defined.

Limitations of our study include the fact that we were restricted by the outcome measures used in the original studies. We examined only WOMAC function, SF-36 physical function, and ODQ, and the latter 2 measures were used in just 1 study each. It is important to examine a wider range of outcome measures, especially those that contain many items for each domain including I, and are likely to measure distinctive outcomes, such as the disease-specific Oxford knee and Oxford hip scores. It will also be important to examine studies with longer followup. We examined data from an RCT comparing home exercise with no intervention in men and women aged  $\geq 45$  years with knee pain<sup>37</sup>, in which the first assessment took place at 6 months, and no differential effect was found. Another limitation may be that in all the studies examined, there was either no effect or only a very small effect for the intervention. Studies with larger effect would perhaps help to better detect how different interventions might differentially affect health domains.

## ACKNOWLEDGMENT

We thank Prof. Elaine Hay and Dr. Elaine Thomas for commenting on an early draft and providing the raw data of the TOPIK trial. We also thank Dr. Kim Thomas and Dr. Isabel Reading for providing data for KIVIS, the WEST, and the knee pain trials. Thanks also to the "Mobile" team (Bristol) for comments and suggestions at early presentations of relevance to this study.

## REFERENCES

1. World Health Organization. International classification of functioning, disability and health: ICF. Geneva: World Health Organization; 2001.
2. Jette AM, Tao W, Haley SM. Blending activity and participation sub-domains of the ICF. *Disabil Rehabil* 2007;29:1742-50.
3. Dawson J, Linsell L, Zondervan K, Rose P, Randall T, Carr A, et al. Epidemiology of hip and knee pain and its impact on overall health status in older adults. *Rheumatology* 2004;43:497-504.
4. Verbrugge LM, Patrick DL. Seven chronic conditions: their impact on US adults' activity levels and use of medical services. *Am J Public Health* 1995;85:173-82.
5. Amlie E, Weber H, Holme I. Treatment of acute low-back pain with piroxicam: results of a double-blind placebo-controlled trial. *Spine* 1987;12:473-6.
6. Weber H, Holme I, Amlie E. The natural course of acute sciatica with nerve root symptoms in a double-blind placebo-controlled trial evaluating the effect of piroxicam. *Spine* 1993;18:1433-8.
7. Bush K, Cowan N, Katz DE, Gishen P. The natural history of sciatica associated with disc pathology. A prospective study with clinical and independent radiologic follow-up. *Spine* 1992;17:1205-12.
8. Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, et al. Linking health-status measurements to the International Classification of Functioning, Disability and Health. *J Rehabil Med* 2002;34:205-10.
9. Weigl M, Cieza A, Harder M, Geyh S, Amann E, Kostanjsek N, et al. Linking osteoarthritis-specific health-status measures to the International Classification of Functioning, Disability, and Health (ICF). *Osteoarthritis Cartilage* 2003;11:519-23.
10. Cieza A, Schwarzkopf S, Sigl T, Stucki G, Melvin J, Stoll T, et al. ICF Core Sets for osteoporosis. *J Rehabil Med* 2004;44 Suppl:81-6.
11. Dreinhofer K, Stucki G, Ewert T, Huber E, Ebenbichler G, Gutenbrunner C, et al. ICF Core Sets for osteoarthritis. *J Rehabil Med* 2004;44 Suppl:75-80.
12. Cieza A, Stucki G, Weigl M, Disler P, Jackel W, van der LS, et al. ICF Core Sets for low back pain. *J Rehabil Med* 2004;44 Suppl:69-74.
13. Schwarzkopf SR, Ewert T, Dreinhofer KE, Cieza A, Stucki G. Towards an ICF Core Set for chronic musculoskeletal conditions: commonalities across ICF Core Sets for osteoarthritis, rheumatoid arthritis, osteoporosis, low back pain and chronic widespread pain. *Clin Rheumatol* 2008;27:1355-61.
14. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. 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.
15. 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.
16. Ware JE Jr, Snow KK, Kosinski SM, Gandek B. SF-36 Health Survey; manual and interpretation guide. Boston: New England Medical Center, Health Institute; 1993.
17. The Oswestry Disability Index (ODI) Version 2.0 or Oswestry Low Back Pain Disability Questionnaire. [Internet. Accessed May 10, 2010.] Available from: [www.medicalcriteria.com/site/index.php?option=com\\_content&view=article&id=171%3Aneurooswestry&catid=64%3Aneurology&Itemid=80&lang=en](http://www.medicalcriteria.com/site/index.php?option=com_content&view=article&id=171%3Aneurooswestry&catid=64%3Aneurology&Itemid=80&lang=en)
18. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271-3.
19. Arden NK, Reading IC, Jordan KM, Thomas L, Platten H, Hassan A, et al. A randomised controlled trial of tidal irrigation vs corticosteroid injection in knee osteoarthritis: the KIVIS Study. *Osteoarthritis Cartilage* 2008;16:733-9.
20. Arden NK, Price C, Reading I, Stubbing J, Hazelgrove J, Dunne C, et al. A multicentre randomized controlled trial of epidural corticosteroid injections for sciatica: the WEST study. *Rheumatology* 2005;44:1399-406.
21. Hay EM, Foster NE, Thomas E, Peat G, Phelan M, Yates HE, et al. Effectiveness of community physiotherapy and enhanced pharmacy review for knee pain in people aged over 55 presenting to primary care: pragmatic randomised trial. *BMJ* 2006;333:995.
22. Hunt SM, McKenna SP. Measuring patients' views of their health. SF 36 misses the mark. *BMJ* 1993;307:125.
23. Ware JE Jr, Keller SD, Hatoum HT, Kong SX. The SF-36 Arthritis-Specific Health Index (ASHI): I. Development and cross-validation of scoring algorithms. *Med Care* 1999;37:MS40-50.
24. Spector WD, Fleishman JA. Combining activities of daily living with instrumental activities of daily living to measure functional disability. *J Gerontol B Psychol Sci Soc Sci* 1998;53:S46-57.
25. Spector WD, Katz S, Murphy JB, Fulton JP. The hierarchical relationship between activities of daily living and instrumental activities of daily living. *J Chronic Dis* 1987;40:481-9.
26. Ayis SA, Bowling A, Gooberman-Hill R, Ebrahim S. The effect of definitions of activities of daily living on estimates of changing ability among older people. *Int J Rehabil Res* 2007;30:39-46.
27. French DJ, Tait RJ. Measurement invariance in the General Health Questionnaire-12 in young Australian adolescents. *Eur Child Adolesc Psychiatry* 2004;13:1-7.
28. Cheung YB. A confirmatory factor analysis of the 12-item General Health Questionnaire among older people. *Int J Geriatr Psychiatry* 2002;17:739-44.

29. Campbell A, Walker J, Farrell G. Confirmatory factor analysis of the GHQ-12: can I see that again? *Aust NZ J Psychiatry* 2003;37:475-83.
30. Everitt BS. *An introduction to latent variable models*. London: Chapman & Hall Ltd.; 1984.
31. Vermunt JK, Magidson J. Latent class models for classification. *Comput Stat Data Anal* 2003;41:531-7.
32. Cieza A, Stucki G. New approaches to understanding the impact of musculoskeletal conditions. *Best Pract Res Clin Rheumatol* 2004;18:141-54.
33. Dreinhofer K, Stucki G, Ewert T, Huber E, Ebenbichler G, Gutenbrunner C, et al. ICF Core Sets for osteoarthritis. *J Rehabil Med* 2004;44 Suppl:75-80.
34. Jette AM, Haley SM, Kooyoomjian JT. Are the ICF Activity and Participation dimensions distinct? *J Rehabil Med* 2003;35:145-9.
35. Badley EM. Enhancing the conceptual clarity of the activity and participation components of the International Classification of Functioning, Disability, and Health. *Soc Sci Med* 2008;66:2335-45.
36. Johnston M, Pollard B. Consequences of disease: testing the WHO International Classification of Impairments, Disabilities and Handicaps (ICIDH) model. *Soc Sci Med* 2001;53:1261-73.
37. Thomas KS, Muir KR, Doherty M, Jones AC, O'Reilly SC, Bassey EJ. Home based exercise programme for knee pain and knee osteoarthritis: randomised controlled trial. *BMJ* 2002;325:752.