

## EXTENDED REPORT

# Inpatient rehabilitation for hip or knee osteoarthritis: 2 year follow up study

M Weigl, F Angst, G Stucki, S Lehmann, A Aeschlimann

*Ann Rheum Dis* 2004;**63**:360–368. doi: 10.1136/ard.2003.011601

See end of article for authors' affiliations

Correspondence to: Professor G Stucki, Department of Physical Medicine and Rehabilitation, University of Munich, Germany; gerold.stucki@phys.med.uni-muenchen.de

Accepted 8 July 2003

**Objective:** To examine the course of pain, physical function, and other health dimensions after a comprehensive inpatient rehabilitation intervention in patients with osteoarthritis (OA) of the hip or knee.

**Methods:** An observational, prospective cohort study with assessments at baseline (entry into clinic), 1 (discharge from inpatient rehabilitation), 3, 6, 9, 12, and 24 months after baseline. Consecutively referred patients to an inpatient rehabilitation centre fulfilling the inclusion criteria were studied. 3–4 week comprehensive rehabilitation intervention, including strengthening exercise, flexibility training, endurance training, relaxation strategies, and consultations for preventive measures, was carried out. Individual home rehabilitation programmes were taught. Generic health status was measured using the SF-36, condition specific health was measured with the WOMAC questionnaire. Effects were analysed with sensitivity statistics (effect size, ES) and non-parametric tests.

**Results:** Data from 128 patients with complete follow up data were analysed. Both pain and physical function improved moderately (WOMAC pain: ES=0.56, WOMAC function ES=0.44) until discharge. Although the effect in pain reduction remained significant by month 24 (WOMAC: ES=0.26), physical function deteriorated close to baseline values after 12 months.

**Conclusions:** Comprehensive inpatient rehabilitation of patients with OA of the hip or knee may improve pain and physical function in the mid-term, and pain in the long term.

Osteoarthritis (OA) is among the five most disabling diseases, having a remarkable public health impact of  $4.7 \times 10^6$  disability adjusted life years (DALYs) in the developed countries in 1990.<sup>1</sup> The prevalence of OA has been assessed at 12% in the American population.<sup>2</sup> Because of its high prevalence in the increasing population of elderly people an increase of  $0.9 \times 10^6$  DALYs (19%) up to  $5.6 \times 10^6$  DALYs is estimated under the present conditions for 2020.<sup>3</sup> Direct and indirect costs for OA of the knee and hip in America in 1994 were \$12.9 billion.<sup>4</sup>

Treatment options for OA are pain relief with analgesics (paracetamol, opioids) and non-steroidal anti-inflammatory drugs (NSAIDs),<sup>5–6</sup> exercise,<sup>7–10</sup> patient education,<sup>11</sup> and finally, joint arthroplasty.<sup>12–13</sup> The recommendations of the European League against Rheumatism<sup>14</sup> and the guidelines of the American College of Rheumatology include all these treatment options.<sup>15–16</sup> Both sets of recommendations emphasise that the treatment should be individually tailored to the patient.

Comprehensive treatment programmes combine some of these treatment options. To the best of our knowledge, no study has analysed the effect size of a comprehensive rehabilitation intervention on patients with OA of the hip or knee in the long term.

This study examined the course of pain and physical function as measured by the condition-specific patient questionnaire, the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC),<sup>17–21</sup> and the generic health status measure, Short Form-36 (SF-36),<sup>22–26</sup> in patients with OA of the hip or knee before a comprehensive rehabilitation intervention and for a 24 month follow up period. We suggested that a comprehensive rehabilitation programme reduces pain and improves function for a period of up to at least 1 year.

## METHODS

### Design

An observational, prospective 24 month cohort study was conducted of patients with OA of the hip or knee who had completed an inpatient rehabilitation intervention.

### Setting

Patients were recruited from the Zurich Rheumatology and Rehabilitation Clinic, Switzerland. They were referred by their family physician or rheumatologist with the diagnosis of hip or knee OA to a comprehensive inpatient rehabilitation intervention. According to the regulations of the Swiss health insurance companies, payment for the inpatient rehabilitation was only granted to patients who still had OA symptoms after completion of four series of nine treatments of outpatient physiotherapy and required further treatment.

### Patients

Consecutive patients with hip or knee OA were invited to participate in the study. Patients were included if they (a) agreed to participate in the study by written, informed consent and (b) fulfilled the American College of Rheumatology criteria for OA (see below). Patients were excluded if they (a) had a history of drug abuse or non-compliance; (b) had not sufficient German language skills to complete the assessment tools; (c) had a severe illness; or (d) did not want to participate in the study. During follow up, patients were excluded if they (a) underwent a joint

**Abbreviations:** DALYs, disability adjusted life years; ES, effect size; MCID, minimal clinically important difference; NSAIDs, non-steroidal anti-inflammatory drugs; OA, osteoarthritis; SDD, smallest detectable difference; SF-36, Short Form-36; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index

arthroplasty; (b) had the following illnesses: terminal cancer, chronic heart failure NYHA IV, asthma requiring continuous oxygen treatment, moderate or severe dementia; (c) died; (d) refused further participation; (e) did not return questionnaires; or (f) if their questionnaires were incomplete according to the missing rules of SF-36 or WOMAC as described below.

**Intervention**

The intervention was a comprehensive rehabilitation intervention of, usually, 3–4 weeks. Its components consisted of individualised physiotherapy, a group exercise programme, group hydrotherapy, swimming, thermal therapy, patient education, massage, and electrotherapy. Table 1 gives details of the interventions. NSAIDs and analgesics were minimised as far as possible. The programme was individually tailored to each patient. Finally, each patient was trained to continue an individual home rehabilitation programme after discharge.

**Measures**

The diagnosis of OA was based on the American College of Rheumatology criteria for OA of the hip or knee.<sup>27, 28</sup> Inclusion criteria for knee OA were (a) knee pain for more than 25 of the past 30 days; (b) morning stiffness of less than 30 minutes; and (c) crepitation in the knee or (a) knee pain for more than 25 of the past 30 days and (b) osteophytes on x ray examination of the knees. Patients with hip OA were included when there was (a) hip pain for more than 25 of the past 30 days and (b) at least two of the following three criteria: (i) erythrocyte sedimentation rate <20 mm/1st h; (ii) osteophytes on x ray examination; or (iii) obliteration of joint space.

Pain and function were measured with the WOMAC<sup>17</sup> and the SF-36.<sup>22</sup> Primary outcome measures were the WOMAC scales for function and pain, and the SF-36 scales for bodily pain and physical functioning. A set of questionnaires including these two instruments was sent or given to the patients at the day of entry (baseline examination), at the day of discharge (about 1 month follow up), and 3, 6, 9, 12, 24 months after baseline.

The WOMAC is an OA-specific multidimensional measure of pain (5 items), stiffness (2 items), and physical functional ability (17 items).<sup>17–20</sup> All 24 WOMAC items are rated on a numerical rating scale ranging from 0 (“no symptoms/no limitation”) to 10 (“maximal symptoms/maximal limitation”), which is the format used in the German validation study.<sup>21</sup> Similar to the visual analogue scale, this rating

provides interval-type data. The mean of the corresponding unweighted item scores results in the score of each scale and, thus also ranges between 0 (no symptoms) and 10 (maximal symptoms). The global WOMAC score was calculated as the unweighted mean of all 24 items.

The SF-36 includes 8 multi-item scales containing 2 to 10 items each plus a single item to assess health transition.<sup>22–24</sup> The scales cover the dimensions of physical functioning (unweighted mean of 10 items), role physical (4 items), bodily pain (2 items), general health (5 items), vitality (4 items), social functioning (2 items), role emotional (3 items), and mental health (5 items), ranging from 0 (“maximal symptoms/maximal limitations/poor health”) to 100 (“no symptoms/no limitations/excellent health”). We used the validated German version.<sup>25, 26</sup> Studies showed excellent psychometric properties and there seems to be a good responsiveness to change in patients with rheumatic conditions, compared with some longer instruments.<sup>29</sup> The SF-36 allows scoring of the eight above mentioned scales and the construction of two summary scales—namely, the physical component summary and the mental component summary scale.

Intake of pain medication was recorded at baseline (entry into the clinic) and at each follow up.

**Data collection procedures**

At entry into the rehabilitation clinic, a trained clinician examined the patient and completed the inclusion and exclusion criteria form. A patient who was included was supplied with and instructed in the completion of self administered questionnaires (WOMAC, SF-36). On the day of discharge, the patients were requested to complete a second set of questionnaires. At 3, 6, 9, 12, and 24 months further questionnaires were sent to the patients for completion.

**Analyses**

Questionnaires were scored in accordance with the missing rules of the WOMAC user’s guide, which specifies completion of at least 4 of the 5 pain items, 1 of the 2 stiffness items, and 14 of the 17 function items.<sup>19</sup> Further, in concordance with the SF-36 Health Survey Manual and Interpretation Guide,<sup>23</sup> each scale of the SF-36 must have at least 50% of the corresponding items answered to be calculable. Patients with in calculable questionnaires either at baseline or at month 24 were excluded.

The responsiveness was assessed by the effect size (ES)<sup>30, 31</sup> at all follow up times. The ES equals the mean change in

**Table 1** Treatment programme

Type of treatment	Content	Goals	Frequency per week (duration)
Individualised physiotherapy*	Specific manual therapy techniques, passive range of motion exercises, assisted range of motion exercises, movement against manual resistance, instruction in home exercise programme	Increase in range of motion, facilitation of movement, reduction in pain, induction of relaxation, muscle strengthening	4 (30 min)
Group exercise programme	Active range of motion exercises, strengthening exercise, weightbearing endurance training, coordination training	Muscle strengthening, improvement of muscle endurance, increase in range of motion, improvement of coordination	2 (30 min)
Group hydrotherapy	Non-weightbearing endurance exercises, strengthening exercises, active range of motion exercises	Muscle strengthening, improvement of endurance, increasing range of motion, induction of relaxation	2 (30 min)
Swimming with flippers	Non-weightbearing muscle strengthening	Muscle strengthening, improvement of endurance	4 (individual)
Thermal therapy	Fango (mud packs)	Induction of muscular relaxation	3 (20 min)
Patient education	Information about disease process, joint protection, pain medication, self management techniques, coping techniques, relaxation techniques, nutrition	Change of behaviour, improvement of self help techniques, improvement in knowledge of treatment options, improvement in the ability to relax	2 (60 min)
Massage†	Deep frictions, classical massage	Induction of relaxation, reduction of pain	Individual
Electrotherapy†	Galvanisation, iontophoresis with diclofenac	Reduction of pain	Individual

\*Intensity and choice of interventions in the individualised physiotherapy were tailored to the patients needs; †massage and electrotherapy were not part of the standard treatment programme. They were prescribed individually to the patients needs.

score divided by the standard deviation of the baseline score. ES thus relates to the change of the mean and to the initial variation in score. An ES of 0.2 is considered a small (beneficial) effect, 0.5 a moderate effect, and 0.8 a large effect of treatment.<sup>30,32</sup> Standard error (SE) and 95% confidence intervals (CIs) for the ES were computed by the SE of "Glass's delta" as described by Rosenthal:<sup>33</sup>

$$\text{SE (of ES)} = \sqrt{(2/n + \text{ES}^2/2(n-1))}$$

where  $n$  = the number of patients. For calculating the CI the following equation was used:

$$95\% \text{ CI} = \text{ES} \pm (1.96 \times \text{SE}).$$

To examine the statistical significance of change in scores, we tested the data for normality and performed paired  $t$  tests if the data were normally distributed, otherwise, we used the Wilcoxon signed ranks test. The null hypothesis was that there was no change. The type I error  $\alpha$  was set at 0.05 and the power at 0.8. All data analysis was performed with SPSS for windows, version 10.0. Follow up scores were illustrated by simple follow up graphs. A SF-36 "spider figure" demonstrated the comprehensive assessment of the health status between two times.<sup>34</sup>

A sensitivity analysis was performed to examine whether the results obtained after 24 months showed any selection bias as a result of drop outs. In studies of older patients with OA and employing a long follow up period, a high drop out rate is to be expected—for example, due to surgery (joint replacement) becoming necessary, change of address, patient death, etc. It is also possible that the patients who drop out of the study have different baseline characteristics from those who remain in the study and complete the follow up. The first step in the sensitivity analysis was a logistic regression analysis to determine which independent variables significantly predicted completion/non-completion of the study with 24 months' follow up (= the dependent variable). For the selection of variables a backward selection was used. The ES was then calculated for given levels (strata) of these predictor variables, using the data available from those who completed the study. Finally, the ES was multiplied by attrition weights, which correct for a change in the distribution of the significant independent variables from baseline to follow up, and which depend on the odds ratios for completion/non-completion.

## RESULTS

### Patients

Three hundred and fifty eight patients were consecutively referred to the rehabilitation clinic between September 1997 and December 1999 with the diagnosis of OA. One hundred and fifty patients did not fulfil the inclusion criteria. From the remaining 208 patients, 80 (38%) had dropped out by month 24 for the following reasons: 11 refused to participate by month 24, 11 did not return the questionnaires or could not be contacted, 48 underwent joint arthroplasty, 6 had a severe illness, 1 died, and 3 could not be analysed owing to missing values, according to the "missing data" rules of the WOMAC and the SF-36. This resulted in 128 patients with complete data.

### Patient characteristics

Table 2 lists the baseline characteristics of all patients

The patients with complete follow up data had a mean age of 63.9 years. Of these patients, 85/128 (66.4%) were female and 84/128 (65.6%) had knee OA. An intake of NSAIDs, analgesics, or both, was reported by 66.4%.

The patients who did not fulfil the inclusion criteria had a mean age 4.9 years greater than that of the patients included ( $t$  test: NS). The patients who underwent a joint arthroplasty during follow up were more often female (79.2%  $v$  66.4%) and took more NSAIDs, analgesics, or both (75.1%  $v$  66.4%), than the patients included. A comparison of the WOMAC pain and function scales (significant group differences only in function,  $t$  test,  $p < 0.05$ ) in the subjects included and excluded showed slightly worse mean scores in those excluded.

The costs of rehabilitation were covered by private health insurance in 15%, by national health insurance in 55%, and by a combination of private and national health insurance in the remaining 30% of cases.

### Responsiveness statistics

Tables 3 and 4 summarise the responsiveness of the rehabilitation at the end of rehabilitation and at 6, 12, and 24 months in comparison with baseline. Testing for normality did not allow performing  $t$  tests for all scales and so we used the Wilcoxon signed ranks test for significance.

The patients showed moderately beneficial effects—that is, ES around 0.5,<sup>33</sup> in the primary outcome variable pain and small to moderate effects in the primary outcome variable physical functioning at the end of rehabilitation. This was measured with the condition-specific WOMAC questionnaire and the generic instrument SF-36 (ES in the WOMAC pain scale was 0.56 and in the SF-36 bodily pain scale 0.52; ES in the WOMAC function scale was 0.44 and in the SF-36 physical functioning scale 0.30).

As opposed to the moderately beneficial effect in pain, there were only small to moderate effects for stiffness (WOMAC stiffness scale ES = 0.38) and small effects in the SF-36 role physical scale (ES = 0.25). The general health scale of the SF-36 showed almost no effect (ES = 0.11).

According to the effects in the subscales, moderate effects at the end of the intervention were also seen in the global WOMAC score (ES = 0.49), which is composed of pain, function, and stiffness, and in the physical component summary of the SF-36 (ES = 0.37), which integrates physical functioning, role physical, bodily pain, and general health.

Still significant, but relatively minor, effects were seen in three of the four SF-36 scales measuring domains of the mental health (vitality ES = 0.25, social functioning ES = 0.23, and mental health ES = 0.19). The SF-36 role emotional scale showed no effect (ES = 0.07).

Figure 1 shows the change of the SF-36 scales between baseline (entry into the clinic) and discharge graphically. Figures 2, 3, and 4 show the course as measured with the WOMAC and SF-36 over 2 years. While the effect in pain reduction remained until month 24 (WOMAC: ES = 0.26; SF-36: ES = 0.49), physical function deteriorated close to the baseline values after 12 months (WOMAC function scale ES = 0.07; SF-36 physical functioning scale ES = 0.07). The already small effects at the end of rehabilitation in the SF-36 scales for vitality, social functioning, and mental health disappeared at the 3 month follow up: vitality ES = 0.19, social functioning ES = 0.13, and mental health ES = 0.11.

### Sensitivity analysis

Logistic regression showed that the following variables significantly predicted completion/non-completion of the study: the affected joint (hip/knee), the baseline scores of WOMAC pain and WOMAC physical functioning, and the SF-36 summary score. Because the WOMAC and the SF-36 both measure health similarly (but the WOMAC does it more responsively) and the two are highly correlated, only the WOMAC scores, in addition to the affected joint, were examined in further detail. For the WOMAC scores ES were

**Table 2** Baseline characteristics of patients with hip or knee OA (n = 358)

Characteristics	Patients with analysable follow up (n = 128)	Excluded at baseline (n = 150)	All drop out patients (n = 80)	Patients dropping out owing to joint replacement (n = 48)
Mean age (years), mean (SD)	63.9 (9.6)	68.8 (11.4)	65.0 (11.3)	64.9 (9.8)
Female (%)	66.4	70.0	68.7	79.2*
Male (%)	33.6	30.0	31.3	20.8*
OA joint				
Knee (%)	65.6	—	37.5*	33.3*
Hip (%)	34.4	—	62.5*	67.7*
Medication at baseline				
NSAID (%)	60.2	—	61.3	66.7
Analgesic (%)	3.9	—	5.1	4.2
NSAID and analgesic (%)	2.3	—	3.8	4.2
Chondroitin sulphate (%)	3.9	—	6.4	4.2
WOMAC pain	4.49	—	4.95	4.91
WOMAC stiffness	4.55	—	5.12*	4.74
WOMAC function	4.30	—	5.29*	5.08
SF-36 PCS	30.3	—	28.4	29.1
SF-36 MCS	49.3	—	46.4	48.2

\*p<0.05.

SD, standard deviation; PCS, physical component scale; MCS, mental component scale from SF-36.

\*Significant differences between patients with analysable follow up and one of the other groups, p<0.05.

then determined for every 10th centile of the baseline scores, to examine where significant changes (“break points”) of the ES occurred. This was the 30th centile for WOMAC pain, and the 50th centile (median) for WOMAC function (data not shown in detail). It was also exactly at these break points that the ES changed in sign—that is, the health status of the “good” patients worsened, and that of the “poor” patients improved. ES above and below these break points were calculated. Table 5 compares the observed ES and the ES that would be expected if the drop out rates in the different strata of WOMAC scores and affected joints were equal. None of the differences between the expected and the observed ES were significant (*t* test, p<0.05).

**DISCUSSION**

In this observational follow up study of patients with hip or knee OA undergoing a comprehensive rehabilitation

intervention, we found a moderate and long term improvement of pain and a moderate but mid-term improvement of physical function.

A sensitivity analysis was performed to account for attrition during the follow up. Baseline pain status, physical function status, and the OA joint were found to have the highest impact on the probability of successfully completing the study. Stratified determination of the effects showed large differences in the ES. For example, the patients who felt little pain (0–30th centile: the “good”) worsened by –0.46 in WOMAC pain, whereas those who reported more pain (31th–100th centile: the “poor”) improved by 0.75. Between baseline and the 24 month follow up, a higher rate of “poor” patients dropped out, so that the expected ES would have been higher (0.38) than the observed ES (0.26) if all patients could have been followed up. That the WOMAC scores of the “poor” are more likely to improve than those of the “good” is

**Table 3** Outcome of patients with hip or knee OA from baseline to the 6 months’ follow up (n = 128)

	Baseline		1 Month			6 Months		
	Mean	(SD)	Mean	(SD)	ES (95% CI)	Mean	(SD)	ES (95% CI)
<b>WOMAC</b>								
Pain	4.49	(2.18)	3.27*	(1.98)	0.56 (0.31 to 0.80)	3.93*	(2.22)	0.26 (0.01 to 0.50)
Stiffness	4.55	(2.61)	3.56*	(2.21)	0.38 (0.13 to 0.63)	4.31	(2.54)	0.09 (–0.15 to 0.34)
Function	4.30	(2.17)	3.34*	(1.99)	0.44 (0.19 to 0.69)	3.96*	(2.18)	0.16 (–0.09 to 0.40)
Global	4.36	(2.10)	3.34*	(1.93)	0.49 (0.23 to 0.74)	4.00*	(2.13)	0.17 (–0.07 to 0.42)
<b>SF-36</b>								
PF	39.8	(20.1)	45.8*	(21.8)	0.30 (0.05 to 0.55)	43.5*	(21.6)	0.18 (–0.06 to 0.43)
RP	24.6	(33.6)	32.9*	(39.9)	0.25 (0.00 to 0.49)	30.8	(38.7)	0.18 (–0.06 to 0.43)
BP	29.9	(17.1)	38.8*	(18.2)	0.52 (0.27 to 0.77)	36.3*	(20.0)	0.37 (0.12 to 0.62)
GH	57.2	(19.9)	59.4	(18.8)	0.11 (–0.13 to 0.36)	55.1	(19.0)	–0.11 (–0.35 to 0.14)
VT	45.9	(19.4)	50.8*	(19.4)	0.25 (0.01 to 0.50)	45.7	(21.3)	–0.01 (–0.26 to 0.23)
SF	69.4	(25.6)	75.2*	(24.6)	0.23 (–0.02 to 0.47)	67.9	(25.5)	–0.06 (–0.30 to 0.19)
RE	53.9	(44.5)	56.8	(45.5)	0.07 (–0.18 to 0.31)	54.6	(46.5)	0.02 (–0.23 to 0.26)
MH	66.3	(18.8)	69.8*	(17.9)	0.19 (–0.06 to 0.43)	65.0	(20.4)	–0.07 (–0.31 to 0.18)
PCS	30.3	(8.3)	33.4*	(8.5)	0.37 (0.12 to 0.62)	32.4*	(9.1)	0.25 (0.01 to 0.50)
MCS	49.2	(11.9)	50.7	(11.5)	0.13 (–0.12 to 0.37)	48.1	(12.9)	–0.09 (–0.34 to –0.15)

\*p<0.05.

ES, effect size (score at follow up – score at baseline)/standard deviation at baseline; PF, physical functioning; RP, role physical; BP, bodily pain; GH, general health; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health; PCS, physical component scale; MCS, mental component scale.

WOMAC scales: 0=no symptoms, 10=maximal symptoms; SF-36 scales: 0=worst health, 100=best health.

Wilcoxon signed ranks test, p one-tailed, null hypothesis: difference (mean at follow up minus mean at baseline = entry into the clinic) = zero; p>0.05: we accept the null hypothesis at type I error = 5%—for example, there is no difference.

The equation for the CI of the ES is not equivalent to the Wilcoxon signed rank test. Therefore in a few cases the CI contains 0, although the Wilcoxon signed rank test is significant.

**Table 4** Outcome of patients with hip or knee OA from 12 months' to 24 months' follow up (n = 128)

	12 Months			24 Months		
	Mean	(SD)	ES (95% CI)	Mean	(SD)	ES (95% CI)
<b>WOMAC</b>						
Pain	3.95*	(2.38)	0.25 (0.00 to 0.49)	3.93*	(2.39)	0.26 (0.01 to 0.50)
Stiffness	4.28	(2.41)	0.10 (-0.14 to 0.35)	4.37	(2.49)	0.07 (-0.18 to 0.31)
Function	4.14	(2.31)	0.07 (-0.17 to 0.32)	4.31	(2.36)	0.00 (-0.25 to 0.24)
Global	4.11	(2.27)	0.12 (-0.13 to 0.36)	4.23	(2.31)	0.06 (-0.18 to 0.31)
<b>SF-36</b>						
PF	41.3	(23.3)	0.07 (-0.17 to 0.32)	42.0	(24.8)	0.11 (-0.14 to 0.35)
RP	29.9	(38.6)	0.16 (-0.09 to 0.40)	35.7*	(41.1)	0.33 (0.08 to 0.58)
BP	38.5*	(20.9)	0.50 (0.25 to 0.76)	38.3*	(22.0)	0.49 (0.24 to 0.74)
GH	56.5	(20.6)	-0.04 (-0.28 to 0.21)	54.8	(19.5)	-0.12 (-0.19 to -0.30)
VT	46.1	(20.9)	0.01 (-0.23 to 0.26)	47.0	(22.1)	0.06 (-0.32 to 0.17)
SF	68.2	(26.8)	-0.05 (-0.29 to 0.20)	67.4	(27.4)	-0.08 (-0.32 to 0.17)
RE	54.6	(45.0)	0.02 (-0.23 to 0.26)	57.3	(47.1)	0.08 (0.17 to 0.32)
MH	65.7	(18.8)	-0.03 (-0.28 to 0.21)	64.4	(20.5)	-0.10 (-0.35 to -0.14)
PCS	32.3*	(9.6)	0.24 (-0.01 to 0.49)	32.8*	(10.6)	0.30 (0.05 to 0.55)
MCS	48.6	(11.5)	-0.05 (-0.30 to 0.19)	48.1	(12.7)	-0.09 (-0.34 to -0.15)

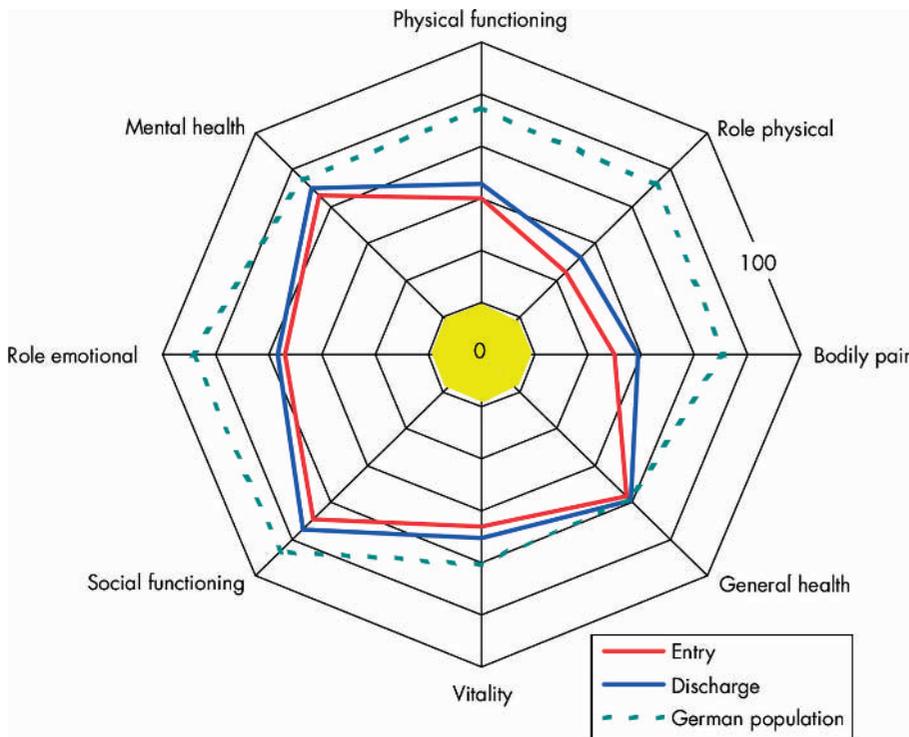
\*p<0.05.  
Explanations are given in the footnotes to table 3.

well known.<sup>35</sup> The effects, therefore, were underestimated. In contrast with this, more (than expected) patients with hip OA—who improved less than the patients with knee OA—dropped out, which led analogously to a slight overestimation (for example, 0.26 v 0.25 in WOMAC pain) of the effect. Our effects, however, would be overestimated, if it were assumed that those patients who showed a negative course—that is, a worsening of health (despite rehabilitation therapy or not), were more likely to undergo joint arthroplasty. In summary, it is difficult to predict what the resulting effects of all the patients would have been if they could all have been followed up; however, this is the situation we meet daily in clinical practice. It is, nevertheless, worth examining the effects in

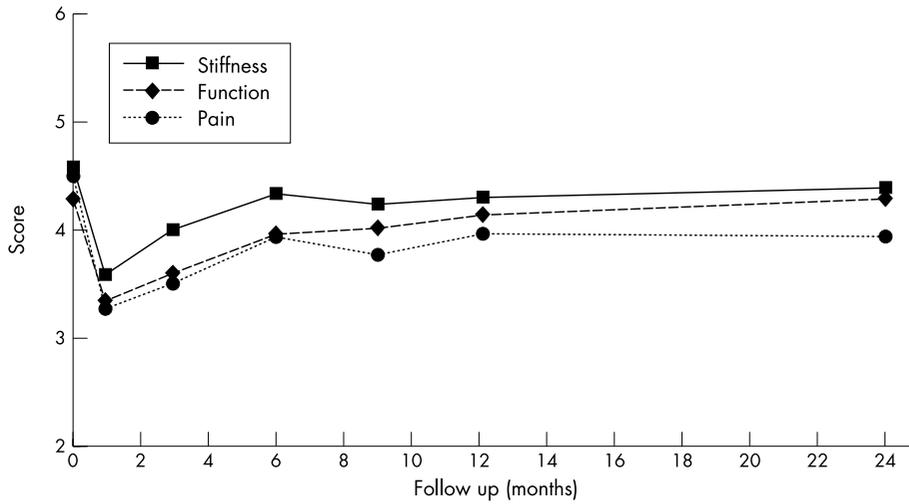
certain subsamples to understand more about the course and the treatment effects.

In this study we applied an observational, naturalistic design within a routine rehabilitation setting with “real” patients. According to the limitations of this design, as described above, the true ES may vary to some extent from the reported treatment effects. On the other hand, this design provides the possibility of obtaining some evidence that exercise therapy improves function and pain not only under the specific conditions of a randomised controlled trial but also in the clinical reality of a common rehabilitation setting.

The ES at the end of rehabilitation were in the same range as in clinical trials with exercise therapy. In a systemic review



**Figure 1** Global health status of patients with hip and knee OA by the SF-36 scales at baseline and at discharge (n = 128). Baseline=entry into the clinic. Discharge of the clinic=about 1 month after baseline. A larger “spider” figure represents a better health status. 0=worst health; 100=best health.



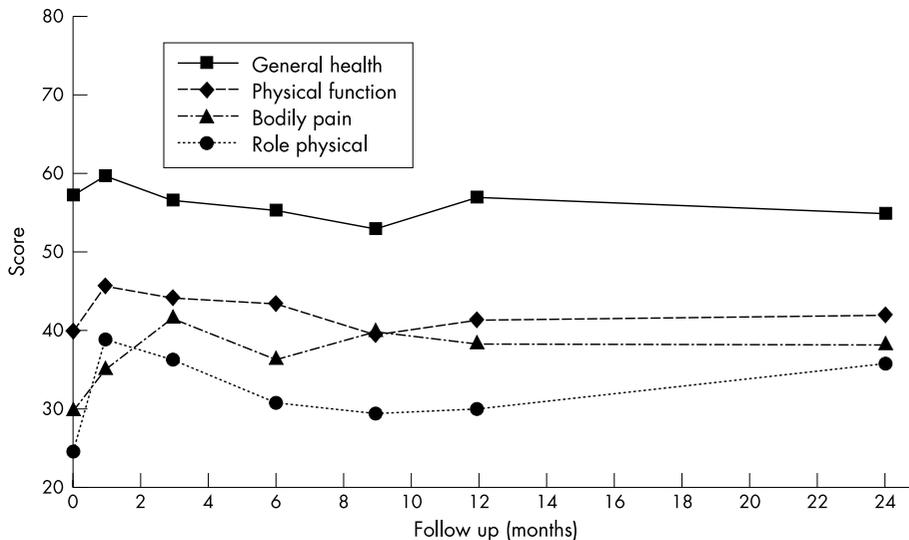
**Figure 2** A 2 year follow up of patients with hip and knee OA by WOMAC scales (n = 128). Patients were assessed with the WOMAC at baseline (= entry into the clinic), at discharge (1 month after baseline), and at months 3, 6, 9, 12, and 24. 0 = no symptoms; 10 = maximal symptoms.

in 1999 van Baar *et al* found three OA exercise therapy studies with good validity.<sup>36</sup> The first of these studies<sup>8</sup> showed an ES of 0.58 for pain measured with a visual analogue scale, and an ES of 0.26 for self reported disability measured with the IRGL questionnaire (Influence of Rheumatic Disease on General Health and Lifestyle) after 12 weeks of treatment. The second and third study,<sup>7</sup> which were performed in one setting (FAST—the Fitness Arthritis and Seniors Trial), used the knee pain scale as measure for pain and a specially developed questionnaire for self reported physical disability. After 18 months, the ES of pain were 0.47 and 0.31 and of self reported disability the ES were 0.41 and 0.36. Although the magnitude of the ES was influenced by the responsiveness of the outcome instruments, it seems likely that the 4 week comprehensive inpatient intervention in this study had a comparable effect to these three exercise interventions.

Comprehensive inpatient rehabilitation programmes are usually more expensive than outpatient exercise programmes, which seem to be similarly effective. However, for some patients there may be factors which indicate that inpatient treatment is preferable. For example, comorbidities like coronary heart disease, which often lead to exclusion from outpatient treatment due to the risk of complications. For instance, in the "Fitness Arthritis and Seniors Trial",<sup>7</sup> cardiac ischaemia during an exercise treadmill test was an exclusion criterion. Several other concomitant disorders,

especially pulmonary diseases, obesity, or further problems of locomotion besides OA are common in patients with hip and knee OA. Owing to the available medical facilities, these multiply affected patients can be treated more safely in an inpatient clinic. Additionally, for some patients with coronary heart disease or pulmonary diseases, an operation may be a high risk or even impossible. In those patients, all non-operative treatment options should be used to avoid disability, which would be more expensive than inpatient treatment. Therefore, all main European health insurance companies pay for inpatient treatment if it is likely that disability cannot be prevented by other treatments. Another problematic group for outpatient treatment are elderly patients with severe OA, especially those living in rural areas, who are unable to (or only at high expense) attend for outpatient treatment which requires a large number of sessions. Therefore, the decision for inpatient or outpatient treatment should be made individually according to each patient's situation.

There was no persisting beneficial effect in function measurement in our patients after 1 year, although the patients had been instructed to follow an individual home based exercise programme. This is consistent with the results of a 1 year follow up of patients who participated in a programme of supervised fitness walking and supportive patient education.<sup>37</sup> In that study self reported disability



**Figure 3** A 2 year follow up of patients with hip and knee OA by the SF-36 scales for physical health (n = 128). The times of assessment are listed in the legend to fig 2. 0 = worst health; 100 = best health.

**Table 5** Sensitivity analysis

	Baseline		24 Month follow up		WOMAC pain		WOMAC function	
	Total (n)	Drop out (n)	Observed (n)	Expected (n)	Observed ES	Expected ES	Observed ES	Expected ES
<i>OA joint</i>								
Hip	94	50	44	58	0.18	–	–0.14	–
Knee	114	30	84	70	0.31	–	0.08	–
Total	208	80	128	128	<b>0.26</b>	<b>0.25</b>	<b>0.00</b>	<b>–0.02</b>
<i>WOMAC pain baseline</i>								
Lower 30% (good health)	63	20	43	39	–0.46	–	–	–
Upper 70% (poor health)	145	60	85	89	0.75	–	–	–
Total	208	80	128	128	<b>0.26</b>	<b>0.38</b>	–	–
<i>WOMAC function baseline</i>								
Lower 50% (good health)	104	31	73	64	–	–	–0.59	–
Upper 50% (poor health)	104	49	55	64	–	–	0.69	–
Total	208	80	128	128	–	–	<b>–0.01</b>	<b>0.05</b>

The column “drop out” contains all drop outs until month 24. The expected numbers of patients at the 24 month follow up reflect the numbers of patients that would be observed, if the drop out rates were equal in the different strata. The expected ES reflect the ES that would be observed, if the drop out rates were equal in the different strata. The comparisons between the expected and the observed ES are in bold. No difference between the observed and the expected ES was significant.

measured with the AIMS (Arthritis Impact Measurement scale) improved significantly after an 8 week programme. The patients were encouraged to continue walking. Nevertheless, after 1 year, there was no statistical significant difference in disability between the walking group and the control group. The authors described the long term adherence to the walking as low, but detailed data of the adherence were not published. Low adherence to the instructed home based programme may also be a possible contributor to the disappointing long term outcome in physical functioning in our study.

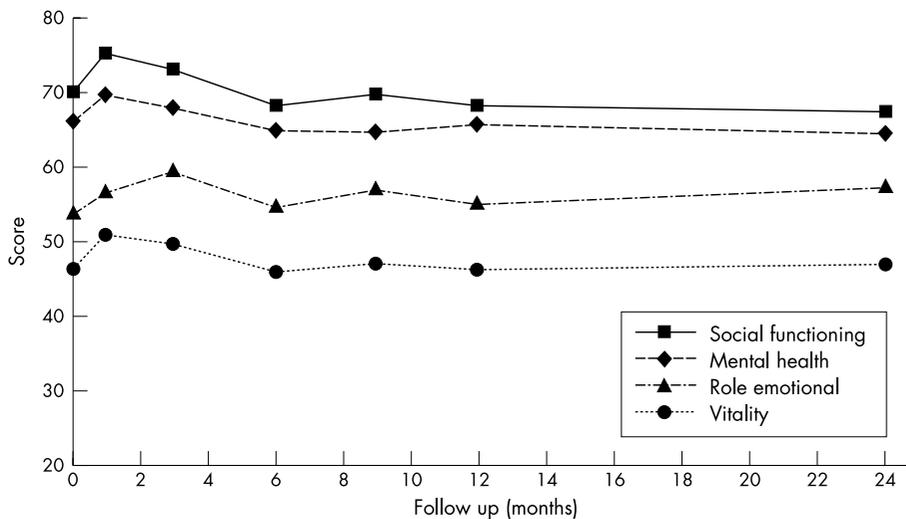
To maintain longlasting functional benefit, exercises probably need to be continued for life. Exercise behaviour before starting an exercise programme has been described as the strongest predictor for compliance.<sup>38</sup> The problem seems to be how to encourage a former sedentary person to do exercises. Short booster sessions, self help groups, or the use of patient diaries have been recommended to increase adherence. However, there is still a lack of controlled clinical trials proving the success of these concepts.

When interpreting the results for physical function, one needs to keep in mind the progressive nature of OA. Therefore, even a stable physical function may represent a

treatment success. However, the progression of illness without treatment was not proved by a control group.

The longlasting pain reduction in this study remains in contrast to the mid-term benefit in physical functioning. This result is in line with a recently published 9 months’ follow up exercise therapy study in patients with OA of the hip or knee.<sup>39</sup> After 24 weeks (12 weeks after completion of treatment) that report found a decreased, but still significant, pain reduction, whereas the improved observed disability had declined to insignificant values.

During follow up, 48 patients underwent joint replacement. This group of patients was defined as dropouts from the long term follow up, because it was impossible to distinguish the effect of the rehabilitation intervention from the effect of the surgical intervention. However, at the end of rehabilitation these patients showed a beneficial ES of 0.57 (compared with 0.49 in the group with complete follow up) in the WOMAC global score. An improved physical function before joint replacement results in an improved outcome after joint replacement.<sup>40</sup> Therefore it can be assumed that this group of patients required less rehabilitation after their joint replacement than they would have needed without the prior rehabilitation intervention.



**Figure 4** A 2 year follow up of patients with hip and knee OA by the SF-36 scales for mental health (n = 128). The times of assessment are listed in the legend to fig 2. 0 = worst health; 100 = best health. The intake of NSAIDs or analgesics or both declined from 66.4% at baseline to 11.0% at the end of rehabilitation. Until month 12 the level of the intake of drugs climbed to the baseline values (67.2%) and did not change significantly until month 24 (63.3%).

The patients with hip OA more often underwent joint arthroplasty during follow up than the patients with knee OA (34% v 14%), although the WOMAC global scores at baseline were higher in the patients with knee OA (knee 4.9 v hip 4.5). We suggest that the difference in the rate of joint arthroplasty is because knee arthroplasty has been developed more recently, and the recovery in pain and function is slower in patients receiving knee arthroplasty than in patients with hip arthroplasty.<sup>41</sup> Therefore, the patients as well as the orthopaedic surgeons may be more predisposed towards hip arthroplasties.

The patients who underwent total joint arthroplasty in this study were younger (mean age 64.9 v 68.8–71.2 years) and had lower WOMAC global baseline scores (4.36 v 5.61–5.98) than patients in other clinical studies in Europe.<sup>42–43</sup> The earlier point in time of operation may be a result of the education of treatment options during the rehabilitation intervention. This may be a benefit because patients operated on earlier in the course of functional decline may have better outcomes.<sup>44</sup>

Patients referred to the clinic with OA but not included in the study were 4.9 years older than the patients included. This difference may have been caused by exclusion criteria such as “difficulties in completing questionnaires”, “having severe illness”, or “joint arthroplasty planned within the next month”. Accordingly, the health status of the excluded patients might have deteriorated, which could limit the generalisability of this study.

Before implementing a treatment, the clinician wants to know if a statistically significant effect is clinically meaningful. Therefore, the concept of minimal clinically important difference (MCID) can be helpful. The MCID is the minimal effect that patients consider clinically perceptible. For the WOMAC pain scale, an improvement of 18% of the mean baseline value was described as the MCID.<sup>35–45</sup> In our study 18% of the mean baseline value was 0.81 points. The minimal clinical important ES in this study can be calculated as follows:  $ES = 0.81/SD \text{ (of mean baseline)} = 0.37$ . Up to the 3 months' follow up, the ES was greater than 0.37. At the later follow up the effect measured with the WOMAC pain scale remained statistically significant, but the ES was smaller than 0.37 and, therefore, these effects may not be clinically meaningful.

When planning a study it is important to know if an MCID can be detected by the chosen outcome instruments. Therefore the smallest detectable difference (SDD) should be calculated.<sup>45</sup> In this study the SDD for improvement of pain measured with the WOMAC was an ES of 0.35 (power = 0.8, type I error = 0.05). The SDD in this study was smaller than the MCID. Thus, the WOMAC pain scale was sensitive enough to detect the MCID in pain.

As before, when we measured 223 patients with OA, including the sample in this study, for 3 months we found that the WOMAC was more responsive than the SF-36 to the improvement in physical function and similarly responsive to pain reduction at the end of rehabilitation.<sup>46</sup> At the 12 month and 24 month follow up, the SF-36 was more responsive to pain. These results indicate that the WOMAC seems to be the better instrument for measuring self reported disability after rehabilitation, whereas the SF-36 bodily pain scale more adequately detects a long term pain benefit. This should be considered in the choice of outcome instruments in future clinical trials.

In conclusion, this study suggests that comprehensive inpatient rehabilitation of patients with hip or knee OA can result in a substantial long term reduction of pain, which declines only slightly with time, and in a mid-term benefit in physical function. Future follow up studies should examine the adherence to instructed exercise programmes, and when

randomised could answer the question whether comprehensive rehabilitation programmes have a preventive effect by stabilising physical function. The encouraging long term benefit in pain in this study should be considered when discussing the cost effectiveness of comprehensive rehabilitation programmes.

## ACKNOWLEDGEMENTS

This study was supported by the Zurzach Spa Foundation, Switzerland.

We thank Susanne Lehmann for the management of the patients, the questionnaires, and the database, and Isabelle Ginsberg and Assistant Professor Anne F Mannion, for the English editing.

## Authors' affiliations

**M Weigl, G Stucki**, Department of Physical Medicine and Rehabilitation, University of Munich, Germany

**F Angst, S Lehmann, A Aeschlimann**, Rehaclinic Zurzach, Switzerland

## REFERENCES

- Murray CJ, Lopez AD, eds. *The global burden of disease; a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Cambridge Harvard University Press, 1996.
- Lawrence RC, Kelsey JL, McDuffie FC, Medsger TA Jr, Felts WR, Shulman LE. Estimates of the prevalence of selected arthritic and musculoskeletal diseases in the United States. *J Rheumatol* 1989;16:427–41.
- Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. *Lancet* 1997;349:1498–504.
- Yelin E. The economics of osteoarthritis. In: Brandt K, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York: Oxford University Press, 1998:23–30.
- Bellamy N, Chalmers A, Ford PM, Kean WF, Kraag GR, Gerez-Simon E, et al. A multicenter study of tenoxicam and diclofenac in patients with osteoarthritis of the knee. *J Rheumatol* 1993;20:999–1004.
- Dieppe P, Jasani MK, McCrae F, Watt I. A two-year, placebo-controlled trial of non-steroidal anti-inflammatory therapy in osteoarthritis of the knee joint. *Br J Rheumatol* 1993;32:595–600.
- Fisher NM, Abrams M, Hicks J, Horrigan D, Pendergast DR. Quantitative effects of physical therapy on muscular and functional performance in subjects with osteoarthritis of the knees. *Arch Phys Med Rehabil* 1993;74:840–7.
- Etinger WH, Burns R, Messier SP, Applegate W, Rejeski WJ, Morgan T, et al. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). *JAMA* 1997;277:25–31.
- Van Baar ME, Dekker J, Oostendorp RA, Bijl D, Voorn TB, Lemmens JA, et al. The effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a randomized controlled trial. *J Rheumatol* 1998;25:2432–9.
- Deyle GD, Henderson NE, Matekel RL, Ryder MG, Garber MB, Allison SC. Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee. *Ann Intern Med* 2000;132:173–81.
- Maurer BT, Stern AG, Kinossian B, Cook KD. Osteoarthritis of the knee: isokinetic quadriceps exercise versus an educational intervention. *Arch Phys Med Rehabil* 1999;80:1293–9.
- Anderson JG, Tsai D, Stulberg SD, Chang RW. Functional outcome and patient satisfaction in total knee patients over the age of 75. *J Arthroplasty* 1996;11:831–40.
- Coyte PC, Hawker G, Croxford R, Wright JG. Rates of revision knee replacement in Ontario, Canada. *J Bone Joint Surg Am* 1999;81:773–82.
- Pendleton A, Arden N, Dougados M, Doherty M, Bannwarth B, Bijlsma JWJ, et al. EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Ann Rheum Dis* 2000;59:936–44.
- Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, et al. Guidelines for the medical management of osteoarthritis. Part I. Osteoarthritis of the hip. American College of Rheumatology. *Arthritis Rheum* 1995;38:1535–40.
- Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR, et al. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. American College of Rheumatology. *Arthritis Rheum* 1995;38:1541–6.
- 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.
- Bellamy N, Kean WF, Buchanan WW, Gerez-Simon E, Campbell J. Double blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): post validation reapplication of the WOMAC osteoarthritis index. *J Rheumatol* 1992;19:153–9.
- Bellamy N. *WOMAC osteoarthritis index. A user's guide*. London, Ontario, Canada: University of Western Ontario, 1995.

- 20 **Bellamy N**, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al*. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus Development at OMERACT III. *J Rheumatol* 1997;**24**:799–802.
- 21 **Stucki G**, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, *et al*. Evaluation of a German version of the WOMAC (Western Ontario and McMaster Universities) osteoarthritis index. *Z Rheumatol* 1996;**55**:40–9.
- 22 **Ware JE**, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;**30**:473–83.
- 23 **Ware JE**, Snow KK, Kosinski M, Gandek B. *SF-36 health survey. Manual and interpretation guide*, 2nd ed. Boston, Massachusetts: The National Health Institute, New England Medical Center, 1997.
- 24 **Ware JE**, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) project. *J Clin Epidemiol* 1998;**51**:903–12.
- 25 **Bullinger M**. German translation and psychometric testing of the SF-36 Health Survey: preliminary results from the IQOLA Project. International Quality of Life Assessment. *Soc Sci Med* 1995;**41**:1359–66.
- 26 **Bullinger M**, Kirchberger I. *SF-36 Fragebogen zum Gesundheitszustand*. Hogrefe, Goettingen: Handanweisung, 1998.
- 27 **Altman R**, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al*. The American College of Rheumatology criteria for the classification of osteoarthritis of the hip. *Arthritis Rheum* 1991;**34**:505–14.
- 28 **Altman R**, Asch E, Bloch D, Bole G, Borenstein D, Brand J, *et al*. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. *Arthritis Rheum* 1986;**29**:1039–49.
- 29 **Kantz ME**, Harris WJ, Levitsky K, Ware JE Jr, Davies AR. Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. *Med Care* 1992;**30**(suppl 5):MS240–52.
- 30 **Kazis ES**, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;**27**(suppl 3):S178–89.
- 31 **Wright JG**, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol* 1997;**50**:239–46.
- 32 **Cohen J**. *Statistical power analysis for the behavioral sciences*. Hillsdale (NJ): Lawrence Erlbaum Associates, 1988.
- 33 **Rosenthal R**. Parametric measures of effect size. In: Cooper H, Hedges HV, eds. *The handbook of research synthesis*. New York: Sage Foundation 1994;231–44.
- 34 **Steiner WA**, Aeschlimann AG. An alternative way to display SF-36 results for individual patients. *Phys Ther* 1998;**78**:1118–19.
- 35 **Angst F**, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *J Rheumatol* 2002;**29**:131–8.
- 36 **Van Baar ME**, Assendelft W J, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee. A systematic review of randomized controlled trials. *Arthritis Rheum* 1999;**42**:1361–9.
- 37 **Sullivan T**, Allegrante JP, Peterson MGE, Kovar PA, MacKenzie CR. One-year follow-up of patients with osteoarthritis of the knee who participated in a program of supervised fitness walking and supportive patient education. *Arthritis Care Res* 1998;**11**:228–33.
- 38 **Rejeski WJ**, Brawley LR, Ettinger W, Morgan T, Thompson C. Compliance to exercise therapy in older participants with knee osteoarthritis: implications for treating disability. *Med Sci Sports Exerc* 1997;**29**:977–85.
- 39 **Van Baar ME**, Dekker J, Oostendorp RA, Bijl D, Voorn TB, Lemmens JA, *et al*. Effectiveness of exercise in patients with osteoarthritis of hip or knee: nine months' follow up. *Ann Rheum Dis* 2001;**60**:1123–30.
- 40 **Parent P**, Moffet H. Preoperative predictors of locomotor ability two months after total knee arthroplasty for severe osteoarthritis. *Arthritis Rheum* 2003;**49**:36–50.
- 41 **Salmon P**, Hall GM, Peerbhoy D, Shenkin A, Parker C. Recovery from hip and knee arthroplasty: patients' perspective on pain, function, quality of life, and well-being up to 6 month postoperatively. *Arch Phys Med Rehabil* 2001;**82**:360–6.
- 42 **Theiler R**, Sangha O, Schaeren S, Michel BA, Tyndall A, Dick W, *et al*. Superior responsiveness of the pain and function sections of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) as compared to the Lequesne-algofunctional Index in patients with osteoarthritis of the lower extremities. *Osteoarthritis Cartilage* 1999;**7**:515–19.
- 43 **Nilsdotter AK**, Aurell Y, Siösteen AK, Lohmander LS, Roos HP. Radiographic stage of osteoarthritis or sex of the patient does not predict one year outcome after total hip arthroplasty. *Ann Rheum Dis* 2001;**60**:228–32.
- 44 **Fortin PR**, Penrod JR, Clarke AE, St-Pierre Y, Joseph L, Belisle P, *et al*. Timing of joint replacement affects clinical outcome among patients with osteoarthritis of the hip or knee. *Arthritis Rheum* 2002;**46**:3327–30.
- 45 **Angst F**, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Rheum* 2001;**45**:384–91.
- 46 **Angst F**, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. *Ann Rheum Dis* 2001;**60**:834–40.