What's all that noise?: The effect of co-morbidity on health outcome questionnaire results after knee arthroplasty

Dunbar, Michael; Robertsson, Otto; Ryd, Leif

Published in:
Acta Orthopaedica Scandinavica

DOI:
10.1080/00016470412331294355

2004

What's all that noise?

The effect of co-morbidity on health outcome questionnaire results after knee arthroplasty

Michael J Dunbar[1,2], Otto Robertsson[1] and Leif Ryd[1]

[1]Department of Orthopaedics, Lund University Hospital, SE-221 85 Lund, Sweden, [2]Division of Orthopaedics, QE II Health Sciences Centre, Dalhousie University, Halifax, Nova Scotia, B3M 1N5, Canada.
Correspondence MJD: Michael.dunbar@ort.lu.se
Submitted 03-02-13. Accepted 03-10-22

Background We modified the Charnley Classification for hips to facilitate its use with knee arthroplasty patients and investigated what affect the different classes of co-morbidity had on the results of a spectrum of outcome questionnaires.

Patients and methods 3600 patients from the Swedish Knee Arthroplasty Registry were surveyed by post with a variety of questionnaires ranging from multiple-item general health, to a single-item knee arthroplasty specific questionnaire. All patients also completed a co-morbidity questionnaire, from which a modified Charnley Classification was generated for each patient. We then investigated the correlation and relationship between the results of the questionnaires and the different classes of co-morbidity.

Results The results of the questionnaires tested varied significantly by Charnley Class, regardless of the specificity of the questionnaire used.

Interpretation We suggest that co-morbidity should be taken into account in outcome studies utilizing general health or disease/site specific questionnaires.

The Orthopedic community is increasingly relying on health outcome questionnaires to define and contrast the value of joint replacement surgery. However, questionnaires are imperfect and their results can be confounded by noise from sources other than the signal of interest (Ryd et al. 1997). Sources of noise include age, gender, pre-operative diagnosis, and co-morbidity. Without recognizing and controlling for the sources of noise, the value of questionnaires for assessing outcomes after arthroplasty is suspect (Gross 1988).

Charnley (1979) recognized the importance of accounting for co-morbidity when assessing outcomes after hip arthroplasty and advocated stratifying patients by degree of co-morbidity to allow for meaningful comparison. The resulting patient strata represent a functional classification and are often referred to as the "Charnley Class". Previously studies have found that the results of health outcome questionnaires applied to hip arthroplasty patients were significantly influenced by Charnley Class (Garellick et al. 1998).

The effect of Charnley Class, or co-morbidity, on the results of health outcome questionnaires applied to knee arthroplasty patients has not been well-defined (Dawson et al. 1996b, Brinker et al. 1997, Hawker et al. 1998). Therefore, the purpose of this study was to first modify the Charnley Classification for application to knee arthroplasty patients and then determine what effect co-morbidity, as defined by the modified Charnley Class, had on the results of a spectrum of outcome questionnaires. Our hypothesis was that general health questionnaires would be influenced by modified Charnley Class, disease specific questionnaires less so, joint specific questionnaires minimally, and a single-item questionnaire about the index knee not at all.

Patients and methods

3600 patients were randomly selected from the Swedish Knee Arthroplasty Register to be mailed a combination of questionnaires. The inclusion criteria were patients with primary osteoarthrosis, age 55–95 at the time of surgery, medial uni-compartmental, lateral uni-compartmental, bilateral (same knee) uni-compartmental or total knee arthroplasty. Patients having undergone an extraction arthroplasty, amputation or arthrodesis were excluded. Patients with bilateral knee arthroplasties were randomly indexed to either their right or left knee so that they were not included twice in the study.

The 3600 patients selected were sent two single-item questionnaires. The first asked the patient to rate their impression of how their index knee felt on a scale of 1–10 (Single-Item Knee Score), and the asked the patient to rate the impression of their general health on a scale of 1–10 (Single-Item Health Score). For both questionnaires a score of 1 represented the worst possible score and a score of 10 represented the best. Of the 3600 patients, 900 were also sent the SF-36, 1200 were sent the WOMAC and 1200 were sent the Oxford-12 Item Knee Score. Additional questionnaires were sent to the remaining patients as part of a separate study and the results from these questionnaires have recently been published (Dunbar et al. 2001).

The SF-36 is a general health questionnaire with 36 questions that yield 8 domain scores and 2 summary scores—Physical Component Summary and Mental Component Summary (Brazier et al. 1992, Ware et al. 1993). We used summary scores standardized (t-transformed) such that the mean score was 50 with scores less than 50 representing a better perception of health than scores above 50 (Brazier et al. 1992, Ware et al. 1993). We used the Swedish Standard Version-I (Sullivan et al. 1995).

The WOMAC (Western Ontario and McMaster Universities Osteoarthritis Index) is a disease specific questionaire and consists of 24 questions that are arranged into 3 domains—Pain, Stiffness, and Physical Function (Bellamy et al. 1988). Pain scores range from 0 to 20, Stiffness scores from 0 to 8, and Physical Function scores from 0 to 68. In all 3 domains, a score of zero represents no problems in the respective domain. The Swedish Likert Version 3.0 was used (Roos et al. 1999).

The Oxford-12 Item Knee Score is a joint specific questionnaire designed specifically for knee arthroplasty (Dawson et al. 1998). 12 Likert-type questions result in a score of 12–60, with 12 representing no perceived problems. We used the Swedish Version 1.0 (Dunbar et al. 2000).

Finally, a three-part questionnaire was sent to all 3600 patients inquiring whether they had arthrosis or an arthroplasty in their contralateral knee, whether they had remote arthrosis affecting their ability to ambulate (e.g., hips, spine, feet), and whether they had a significant medical condition affecting their ability to ambulate (e.g., congestive heart failure, angina, pulmonary disease, cerebral vascular accident, etc.). These questions were used to formulate a modified Charnley Class for knees to account for co-morbidity. Class A referred to patients with disease in the index operated knee only. Class B1 referred to patients with one knee arthroplasty and untreated arthrosis (no arthroplasty) in the contra lateral knee. Class B2 referred to patients with bilateral arthroplasties and Class C referred to patients with a knee arthroplasty and remote arthritis and/or a medical condition that affected their ability to ambulate.

A postage paid return envelope was included with the questionnaires. Reminder letters were sent at 2 weeks. An individual patient's questionnaires was included in the analysis only if they responded to all the questions within (complete questionnaire) and if they completed the 3-part Charnley questionnaire on co-morbidity.

The average patient age at the time of mail-out was 78 years (SD 7.1, range 58–94) and 71 years (SD 6.7, range 55–90) at the time of index surgery. The average follow-up time was 6.7 years (SD 3.8, range 1.4–21). 70% (n = 2511) of the sample were female. 95% (n = 3402) were primary arthroplasties. 58% (n = 2084) of all patients had tri-compartmental knee replacements, 36% (n = 1296) had medial uni-compartmental knee replacements, and 6.1% (n = 220) had either a lateral uni-compartmental or both compartments of the same knee replaced with an uni-compartmental prosthesis.

We performed multinomial regression to determine what effect gender, patient age and year of operation had on the modified Charnley Class. ANOVA was used to compare mean ages between

Table 1. Age distributions for Charnley Class by gender

Charnley Class	Women			Men		
	n	mean	95%CI	n	mean	95%CI
A	289	77	76–78	212	77	76–77
B1	188	76	75–77	112	75	74–76
B2	284	78	77–78	142	77	76–78
C	1269	78	78–78	437	77	76–77
All classes	2030	78	77–78	903	76	76–77
P-value		<0.001			0.06	

Charnley Classes while the Chi Squared test was used to compare the frequency distribution of Charnley Class by gender and by age category (less than 75 years and greater than or equal to 75 years). We determined differences in questionnaire scores by modified Charnley Class with the Kruskal-Wallis test. P-values of less than 0.05 were considered significant. We performed linear regression analyses for each questionnaire with the questionnaire score as the dependent variable and patient age at the time of postal survey, gender, time since operation, type of prosthesis (uni-compartmental versus total), revision status, and modified Charnley Class as the independent variables. We performed logarithmic transformations to normalize the distribution of skewed scores when performing linear regressions (Bland et al. 1996a, b).

Results

Of the 3600 contacted, 0.5% (n = 17) were deceased or unable to be located by the post-office.

Multinomial regression demonstrated that gender and patient age at the time of mail-out affected the modified Charnley Class distribution ($p < 0.001$). ANOVA confirmed the differences in age between Charnley Classes, but the differences were clinically small (maximum difference 2 years) and were only significant for females (Table 1).

The distribution of Charnley Classes differed between females and males ($p < 0.001$) with females having a higher proportion of patients in Charnley Class C even after that age distribution had been accounted for (Figure 1). While there was no difference in the distribution of Charnley

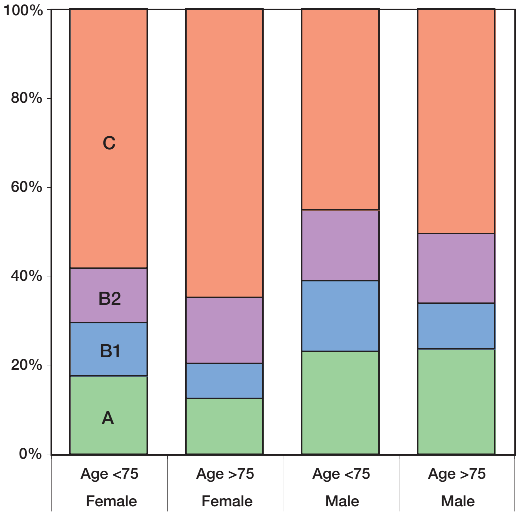

Figure 1. Distribution of modified Charnley Class by age and gender.

Classes between age groups for males, females younger than 75 years had a different distribution compared to those 75 years and older ($p < 0.001$, Figure 1) with older females having a higher frequency of Charnley Class C patients.

For all questionnaires tested, significant differences were found in the scores when analyzed by Charnley Class (Table 2). A consistent pattern emerged for the distribution of scores by Charnley Class (Table 3, Figures 2–4). Patients with mono-articular knee involvement, treated with arthroplasty (Class A) scored the best while patients with one arthroplasty and arthrosis in the contralateral knee (Class B1) scored significantly worse. Patients with bilateral arthroplasties (Class B2) tended to score as if they had no arthrosis in the knee contra lateral to the index knee (i.e., Class A). Patients with knee arthroplasty and remote arthrosis or systemic disease affecting their ability to ambulate (Class C) scored worse than all other classes. These results were found regardless of the type of questionnaire or stratification of scores by gender or patient age (Table 2).

While we noted a consistent pattern in questionnaire scores by Charnley Class, the magnitude of the change varied by questionnaire (Table 4). The WOMAC scores varied the most, with a 75% increase in Physical Function scores when comparing Charnley Class A–B1, and a 138% increase from class A–C. The Oxford-12 scores varied

Table 2. Statistical significance of differences in the score of each questionnaire and its domains when factored by Charnley Class (Kruskal-Wallis test) and age and gender are accounted for

Questionnaire	Women, age < 75	Women, age ≥ 75	Men, age < 75	Men, age ≥ 75
SF-36 Physical component summary	< 0.001	< 0.001	0.008	0.04
SF-36 Mental component summary	0.002	0.05	0.2	0.5
Global health score	< 0.001	< 0.001	< 0.001	< 0.001
WOMAC Pain	< 0.001	< 0.001	0.006	0.1
WOMAC Stiffness	< 0.001	< 0.001	0.01	0.3
WOMAC Physical function	< 0.001	< 0.001	0.001	0.005
Oxford-12 Knee score	< 0.001	< 0.001	0.02	< 0.001
Global knee score	< 0.001	< 0.001	< 0.001	< 0.001

Table 3. Mean score and standard deviation by Charnley Class for health outcome questionnaires (all patients)

Questionnaires	n	Charnley A mean	Charnley A SD	Charnley B1 mean	Charnley B1 SD	Charnley B2 mean	Charnley B2 SD	Charnley C mean	Charnley C SD	All classes mean	All classes SD
SF-36 Physical component summary	484 [a]	57	2.6	65	2.6	62	2.8	70	1.0	67	1.0
SF-36 Mental component summary	484 [a]	49	2.5	48	4.1	47	2.9	54	1.5	52	1.1
Global health score	2736	3.2	0.2	3.8	0.3	3.2	0.2	5.0	0.1	4.3	0.1
WOMAC Pain	934	2.8	0.6	4.6	0.9	3.4	0.6	6.3	0.4	5.1	0.3
WOMAC Stiffness	951	1.3	0.3	2.3	0.4	1.6	0.3	2.8	0.2	2.3	0.1
WOMAC Physical function	836	12	2.3	21	3.4	16	2.5	28	1.5	23	1.1
Oxford-12 Knee score	882	19	1.1	25	1.9	22	1.6	29	1.0	26	0.7
Global knee score	2773	2.9	0.2	3.7	0.3	3.0	0.2	4.4	0.1	3.9	0.1

[a] SF-36 scores have been inverted for comparative purposes

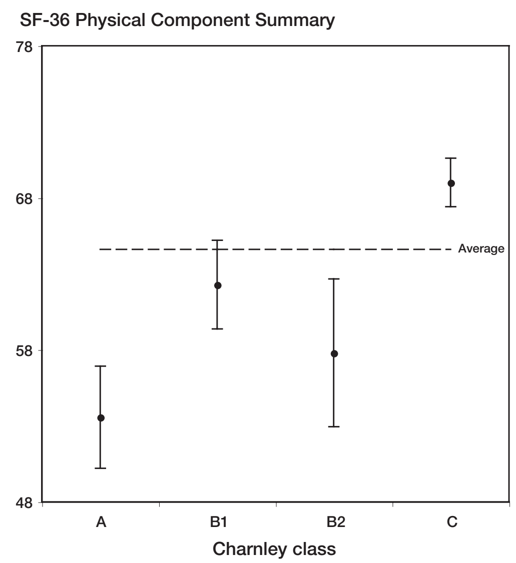

Figure 2. Variation in SF-36 Physical component summary scores by Charnley Class for women < age 75. Error bars represent 95% confidence intervals. Range of scores listed on the Y-axis represent 2 standard deviations. N.B. Scores have been inverted for comparative purposes.

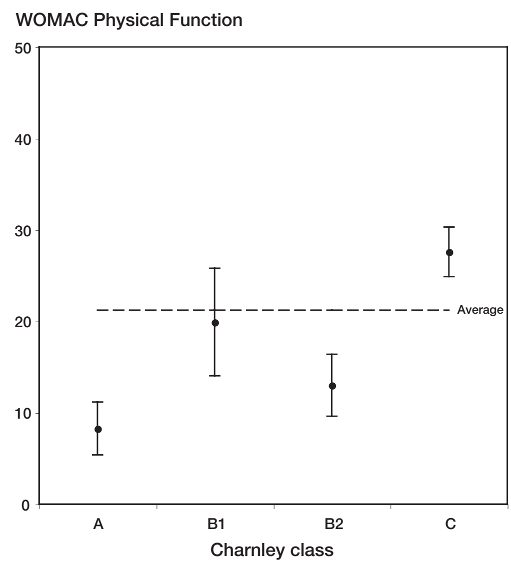

Figure 3. Variation in WOMAC Physical function scores by Charnley Class for women < age 75. Error bars represent 95% confidence intervals. Range of scores listed on the Y-axis represent 2 standard deviations.

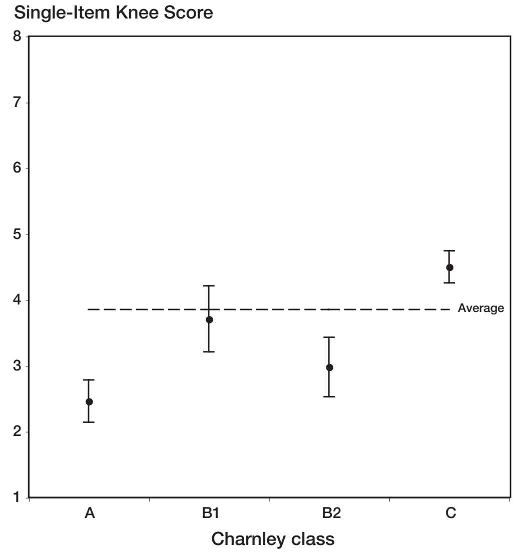

Figure 4. Variation in Single-Item Global Knee scores by Charnley Class for females < age 75. Error bars represent 95% confidence intervals. Range of scores listed on the Y-axis represent 2 standard deviations.

to a lesser degree with a 34% increase in scores between Charnley Class A and B1 and a 55% increase between Charnley Class A and C. Similar changes were noted for the single item Single-Item Knee and Single-Item Health Scores (Table 4). However, the Single-Item Knee and Single-Item Health scores had less of a change between Charnley Class A and B2. The SF-36 Physical and Mental Component Summary scores changed the least by Charnley Class.

Linear regression analyses for the various scores tested demonstrated a variety of covariates as having an effect on the scores, depending on the questionnaire (Table 5). However, for every questionnaire the modified Charnley Class was a significant factor, even when all other factors were accounted for in the regression equation ($p < 0.001$). No other factors were significant for all questionnaires.

Discussion

Patient co-morbidity as stratified by the modified Charnley Classification was a significant factor for all questionnaires tested, regardless of the specificity of the questions to the index knee. This was an unexpected finding. In order to be certain that these result were not a function of different age or sex distributions for modified Charnley Class, we analyzed the data while stratifying by these variables for the Kruskal-Wallis test and by including them along with other covariates in the regression equation. After accounting for all foreseeable sources of error, we still found that Charnley Classes significantly affected the results of questionnaires.

Statistically significant changes in questionnaire scores by Charnley Class do not necessarily imply clinically significant changes. To assess the quantitative impact, we looked at the percentage change in scores by Charnley Class. WOMAC scores more than doubled by Charnley Class while the Oxford-12 and Single-Item Knee and Health scores increased by as much as 55%. Clearly, these changes would be clinically relevant. The SF-36 Physical and Mental Component Summary scores varied to a lesser degree. It is unclear if changes in these scores would be clinically relevant.

Table 4. Percentage change in questionnaire scores by Charnley Class (all patients)

Questionnaire	n	% change in Charnley Class A to B1	A to B2	A to C
SF-36 Physical component summary	484	13	7.7	22
SF-36 Mentall component summary	484	–1.6	–3.9	10
Global health score	2736	20	1.8	58
WOMAC Pain	934	67	23	127
WOMAC Stiffness	951	73	20	107
WOMAC Physical function	836	75	35	138
Oxford-12 knee score	882	34	18	55
Global knee score	2773	29	5.0	54

Table 5. Results of linear regression demonstrating significant factors that affect scores of health outcome questionnaires applied to knee arthroplasty patients

Questionnaire	n	Transf. [a]	Factor	P-value
SF-36 Physical component summary	484	None	Charnley	<0.001
			Age at survey	<0.001
			Gender	0.01
			Type (uni. vs total)	0.03
SF-36 Mental component summary	484	None	Charnley	0.001
Global health score	2736	None	Charnley	<0.001
			Age at survey	<0.001
			Gender	<0.001
			Operative year	0.008
			Revision status	0.05
WOMAC Pain	934	log10	Charnley	<0.001
			Revision status	<0.001
			Gender	0.004
WOMAC Stiffness	951	None	Charnley	<0.001
			Revision status	<0.001
WOMAC Physical function	836	None	Charnley	<0.001
			Revision status	<0.001
			Age at survey	0.02
			Operative year	0.004
			Gender	0.01
			Type (uni. vs total)	0.03
Oxford	882	log10	Charnley	<0.001
			Operative year	<0.001
			Revision status	0.02
			Type (uni. vs total)	0.03
Global knee score	2773	log10	Charnley	<0.001
			Revision s status	<0.001

[a] Transformation required to normalize regression residuals plot.

It could be assumed that the general questions within the SF-36 regarding concepts such as body pain and physical function would be susceptible to the "noise" of co-morbidity when inquiring about the index knee. Hence, the significant differences between Charnley Classes for the SF-36 Physical Component Summary were predictable. The fact that the changes in score by Charnley Class were small and questionably clinically relevant probably refers to the fact that there are no specific questions regarding the knee in the SF-36. Therefore, the signal for knee pathology in this questionnaire can be assumed to be low to begin with.

The disease specific WOMAC questionnaire inquires about pain with activity and the ability to perform activities such as stair climbing, putting on shoes and socks, etc. The noise of remote arthrosis could be expected to impact on the WOMAC scores, as hip or spine arthrosis could cause referred pain and interfere with a patient's ability to complete these tasks. The Oxford-12 score asks more specific questions related to the knee. In this case, less variation in scores by modified Charnley Class could be expected. This could account for the difference in the magnitude of the change in scores. Still, the Oxford-12 score was susceptible to the noise of co-morbidity. However, closer inspection of the Oxford-12 reveals that it too asks questions concerning stair climbing and putting on shoes and socks, hence it too can be rationalized to be susceptible to noise.

In an effort to concentrate singularly on the index knee and to remove any extraneous questions that may pick up on remote arthrosis or systemic disease, we asked all the patients a single question regarding how their index knee felt on a scale of 1–10. Surprisingly, the same pattern occurred as for the other questionnaires and we again found that there were significant differences in this score when compared by Charnley Class. Furthermore, the same magnitude of change in score occurred with this questionnaire as seen with the Oxford-12.

The effects of co-morbidity on surgeon-derived scores for knee arthroplasty patients (e.g., Knee Society Score, Hospital for Special Surgery Knee Score) have been previously investigated (Brinker et al. 1997). Patients having two or more significant medical conditions were found to have worse scores than others without the same level of co-morbidity. Furthermore, the authors concluded that when analyzing groups, without matching for sources of noise, differences in common knee scores between the groups are at least as likely to represent differences in the patient populations as in their treatments (Brinker et al. 1997). This is in general agreement with our findings although our study shows that both remote arthrosis and medical conditions affect patient derived outcome scores.

Garellick et al. (1998) found that the Charnley Class for hips significantly influenced the results of outcome scores applied to hip arthroplasty patients. This too is in agreement with our results for knee arthroplasty patients. Dawson et al. (1996a, b) investigated the effect of remote joint co-morbidity on the change in the SF-36, Arthritis Impact Measurement Scales and the Oxford-12 Item Hip Score from pre and post-operative application. They found that the Oxford-12 Item Hip Score did not detect any difference between groups with and without remote arthrosis, while the other questionnaires did. Based on this, they concluded that the Oxford-12 Item Hip Score was highly joint-specific and was not susceptible to the noise of remote arthrosis. However, it should be emphasized that the differences between these patient groups generated by the remote arthrosis (noise) may have been lost in the profound change in scores seen between pre-operative and post-operative patients (signal), regardless of the co-morbid status (Laupacis et al. 1993, Dawson et al. 1996a). This could explain the discrepancy between our results and theirs, especially since we applied the Oxford-12 Item Knee Score in a discriminative fashion.

The implication of our study is that it is not possible to assess the knee joint with questionnaires in isolation from the rest of the body, but instead, co-morbidity must be accounted for. This is particularly true when patients are evaluated in a discriminative fashion. Without such knowledge, erroneous conclusions could be drawn because of the significant impact that co-morbidity has on questionnaire results. The Charnley Class questionnaire that we employed seems like a convenient and effective way to assess patient co-morbidity when applying outcome questionnaires to knee arthroplasty patients.

In conclusion, we found that co-morbidity has a significant effect on outcome questionnaires after knee arthroplasty, regardless of the specificity of the questionnaire used. Results of questionnaires could vary by as much as 138% between Charnley Classes. Co-morbidity should be accounted for in outcome studies, especially with a discriminative questionnaire application. The modified Charnley Classification questionnaire for knee arthroplasty is a useful method for assessing co-morbidity in this population.

Funding has been provided by grants from the Arthritis Society of Canada, the Medical Research Council of Sweden (Project 9509), Stiftelsen för bistånd åt vanföra I Skåne, the Medical Faculty of the University of Lund, and Socialstyrelsen. The authors wish to thank Jonas Ranstam for statistical consultation.

No competing interests declared.

Bellamy N, Buchanan W W, Goldsmith C H, Campbell J, Stitt L W. 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 (12): 1833-40.

Bland J M, Altman D G. Transformations, means, and confidence intervals. BMJ 1996a; 312 (7038): 1079.

Bland J M, Altman D G. The use of transformation when comparing two means. BMJ 1996b; 312 (7039): 1153.

Brazier J E, Harper R, Jones N M, O'Cathain A, Thomas K J, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992; 305 (6846): 160-4.

Brinker M R, Lund P J, Barrack R L. Demographic biases of scoring instruments for the results of total knee arthroplasty. J Bone Joint Surg (Am) 1997; 79 (6): 858-65.

Charnley J. Low friction arthroplasty of the hip. Springer-Verlag, Berlin 1979.

Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg (Br) 1996a; 78 (2): 185-90.

Dawson J, Fitzpatrick R, Murray D, Carr A. The problem of "noise" in monitoring patient-based outcomes: generic, disease-specific and site-specific instruments for total hip replacement. J Health Serv Res Policy 1996b; 1 (4): 224-31.

Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg (Br) 1998; 80 (1): 63-9.

Dunbar M J, Robertsson O, Ryd L, Lidgren L. Translation and validation of the Oxford-12 item knee score for use in Sweden. Acta Orthop Scand 2000; 71 (3): 268-74.

Dunbar M J, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty. Results of a survey of 3600 patients from The Swedish Knee Arthroplasty Registry. J Bone Joint Surg (Br) 2001; 83 (3): 339-44.

Garellick G, Malchau H, Herberts P. Specific or general health outcome measures in the evaluation of total hip replacement. A comparison between the Harris hip score and the Nottingham Health Profile. J Bone Joint Surg (Br) 1998; 80 (4): 600-6.

Gross M. A critique of the methodologies used in clinical studies of hip-joint arthroplasty published in the English-language orthopaedic literature. J Bone Joint Surg (Am) 1988; 70 (9): 1364-71.

Hawker G, Wright J, Coyte P, Paul J, Dittus R, Croxford R, Katz B, Bombardier C, Heck D, Freund D. Health-related quality of life after knee replacement. J Bone Joint Surg (Am) 1998; 80 (2): 163-73.

Laupacis A, Bourne R, Rorabeck C, Feeny D, Wong C, Tugwell P, Leslie K, Bullas R. The effect of elective total hip replacement on health-related quality of life. J Bone Joint Surg (Am) 1993; 75 (11): 1619-26.

Roos E M, Klassbo M, Lohmander L S. WOMAC osteoarthritis index. Reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. Western Ontario and MacMaster Universities. Scand J Rheumatol 1999; 28 (4): 210-5.

Ryd L, Kärrholm J, Ahlvin P. Knee scoring systems in gonarthrosis. Evaluation of interobserver variability and the envelope of bias. Score Assessment Group. Acta Orthop Scand 1997; 68 (1): 41-5.

Sullivan M, Karlsson J, Ware J E, Jr. The Swedish SF-36 Health Survey. I. Evaluation of data quality, scaling assumptions, reliability and construct validity across general populations in Sweden. Soc Sci Med 1995; 41 (10): 1349-58.

Ware J E, Snow K K, Kosinski M, Gandek B, SF-36 Health Survey: Manual and Interpretation Guide. Nimrod Press, Boston, Massachusetts 1993.