

Predictors of multidimensional outcome after spinal surgery

A. F. Mannion · A. Elfering · R. Staerke ·
A. Junge · D. Grob · J. Dvorak · N. Jacobshagen ·
N. K. Semmer · N. Boos

Received: 30 September 2005 / Revised: 12 July 2006 / Accepted: 23 October 2006 / Published online: 14 November 2006
© Springer-Verlag 2006

Abstract The variables identified as predictors of surgical outcome often differ depending on the specific outcome variable chosen to designate “success”. A short set of multidimensional core outcome measures was recently developed, in which each of the following domains was addressed with a single question and then combined in an index: pain, function, symptom-specific well-being, general well-being (quality of life), disability (work and social). The present study examined the factors that predicted surgical outcome as measured using the multidimensional core measures. 163 spinal surgery patients (mixed indications) completed questionnaires before and 6 months after surgery enquiring about demographics, medical/clinical history, fear-avoidance beliefs (FABQ), depression (Zung self-rated depression), and the core measures domains. Multiple regression analyses were used to identify predictor variables for each core domain and for the multidimensional combined core-set index. The combination of baseline symptoms, medical variables (pain duration, previous spine operations, number of levels treated, operative procedure) and psychosocial factors (FABQ and Zung depression) explained 34% of the variance in the core measures index ($P < 0.001$). With regard to the

individual domain items, the medical variables were better in predicting the items “pain” and “symptom-specific well-being” ($R^2 = 6\text{--}7\%$) than in predicting “function”, “general well-being” or “disability” (each $R^2 < 4\%$). The inverse pattern was shown for the psychosocial predictors, which accounted for in each case approximately 20% variance in “function”, “general well-being” and “disability” but only 12–14% variance in “pain” and “symptom-specific well-being”. Further to previous studies establishing the sensitivity to change of the core-set, we have shown that a large proportion of the variance in its scores after surgery could be predicted by “well-known” medical and psychosocial predictor variables. This substantiates the recommendation for its further use in registry systems, quality management projects, and clinical trials.

Keywords Predictors · Multidimensional · Core outcome measures · Spinal surgery · Psychology

Introduction

It is well known that some patients operated on for spinal disorders will have a poor result, regardless of the apparent technical success of the operative procedure itself [27]. This has prompted the search for risk factors and the development of pre-screening tools [1, 4, 11] to assist with both the patient selection procedure and the promotion of realistic expectations on behalf of the patient. Over the last 10–15 years, numerous studies have sought to identify the determinants of surgical outcome [15]. Despite this, there is still a lack of consensus regarding both the most important predictors and their overall predictive power

A. F. Mannion (✉) · A. Junge · D. Grob ·
J. Dvorak
Schulthess Klinik, Lengghalde 2, Zürich 8008, Switzerland
e-mail: anne.mannion@kws.ch

A. Elfering · N. Jacobshagen · N. K. Semmer
Department of Psychology,
University of Bern, Bern, Switzerland

R. Staerke · N. Boos
Centre for Spinal Surgery,
University of Zürich, Zürich, Switzerland

(i.e. their clinical relevance). The risk factors identified in any given study are most likely contingent upon certain methodological factors—such as the study design (retrospective versus prospective), the statistical methods used (bivariate, multivariate analyses), the number and type of predictors examined, and their prevalence within the patient group examined—as well as factors such as the specific pathology or the surgical procedure being investigated [15]. Furthermore, both the proportion of positive outcomes after spinal surgery [10] and the factors identified as predictors [9, 25] depend to a large extent on the manner in which the outcome itself is assessed.

There is no single, universally accepted method for assessing the outcome of spinal surgery. In the past, many clinicians developed their own simple rating scales, using categories such as “excellent, good, moderate, and poor”, which they themselves used to judge the outcome, predominantly from a surgical or a clinical perspective. The technical success of the operation also lent itself to evaluation in terms of, for example, the accuracy of screw placement or the degree of fusion/extent of decompression achieved, as monitored by appropriate imaging modalities at follow-up. In an effort to achieve further objectivity, these measures were sometimes supplemented with physiologic measures such as range of motion or muscle strength [7]. However, in many cases, these indices proved to be only weakly associated with outcomes of relevance to the patients and to society. With the increasing awareness that the outcome should be (at least also) assessed by the patient himself/herself, the previously popular surgical outcome measures were superseded by a diverse range of patient-orientated questionnaires, assessing factors of importance to the patient, such as symptoms, disability, quality of life, and ability to work. However, the emergence of many new instruments in each of these domains and the lack of their standardized use [34] has compromised meaningful comparison among different diagnostic groups, treatment procedures, and clinical studies [2]. In recognition of this problem, a standardized set of outcome measures for use with back pain patients was proposed in 1998 by a multinational group of experts [7]. There was general consensus that the most appropriate core outcome measures should include the following domains: pain, back-specific function, generic health status (well-being), work disability, and patient satisfaction [2, 7]. Recent studies have shown that these measures, whilst related, are not interchangeable as outcome measures [8]. In 1998, Deyo et al [7] developed a core set of just six questions that would cover all of these domains yet be brief enough to be practical for

routine clinical use, quality management, and possibly also more formal research studies. The psychometric characteristics of the core-set were recently examined in back pain patients undergoing either surgical or conservative treatment [16, 22]. The individual core items as well as a multidimensional sum-score of all the core items showed good reliability, validity, and responsiveness [16, 22].

The present study sought to examine the factors that predict the outcome of spinal surgery as measured using this brief, multidimensional score. Some of the most important demographic/biological, psychosocial, and medical risk factors suggested by a recent review of the literature [15] were included, in order to assess their relative importance to each of the outcome domains and to the multidimensional score as a whole.

Methods

Study population

Patients were recruited (from March 1999 through to April 2000) from the spine centres of two neighbouring orthopaedic hospitals: one was a (public) University Hospital and the other was a not-for-profit Foundation Hospital with University affiliations. The inclusion criteria included fluency in the German language, back pain or leg pain due to a spinal disorder and scheduled for spinal surgery, and willingness to complete a questionnaire booklet before and 6 months after surgery. Exclusion criteria were severe medical problems (e.g. tumour, infection, cardiovascular disease) and spinal problems as a result of trauma. Patients were recruited by research assistants, through direct personal request in the University Hospital, and via a letter in the mail in the Foundation hospital. In the latter hospital, herniated disc patients were not invited to participate, as they were being examined as part of another study. All other consecutive patients who fulfilled the admission criteria were invited to participate; those completing the first questionnaire were sent a second postal questionnaire 6 months after surgery.

Questionnaires

The questionnaire booklet enquired, amongst other things, about the following (for further details, see Table 1):

- Sociodemographic variables (gender and age)
- Clinical/pain history (pain duration, previous operations on the spine)

Table 1 Questionnaires/questionnaire items

Domains/questionnaires	Number of items	Response format	References
Predictor variables			
Pain duration	1	Seven-point Likert scale	
Number of previous operations	1	Four-point Likert scale	
Number of levels treated	1	1–6	
Fear-Avoidance-Beliefs-Questionnaire		Seven-point Likert scale (0–6)	[26, 31]
FABQ1—work beliefs	7		
FABQ2—activity beliefs	4		
Zung Depression Scale	20	Four-point Likert scale (1–4)	[6, 14, 35]
Outcome variables			
(1) Pain symptoms	1	Visual Analogue Scale: 0 (no pain) to 10 (worst pain I can imagine)	Item max of two item values: “How severe was your back pain in the last week?” “How severe was your leg pain in the last week?”
(2) Back function	1	Likert scale 1 (not at all) to 5 (extremely)	“During the past week, how much did your back problem interfere with your normal work (including both work outside the home and housework)?”
(3) Symptom-specific well-being	1	Likert scale 1 (very dissatisfied) to 5 (very satisfied) ^a	“If you had to spend the rest of your life with the symptoms you have right now, how would you feel about it?”
(4) General well-being	1	Likert scale 1 (very bad) to 5 (very good) ^a	“How would you rate your quality of life?”
(5) Disability	1	Number of days	Mean of two item scores: “During the past four weeks, how many days did you cut down on the things you usually do (work, housework, school, recreational activities) because of your back problem?” “During the past four weeks, how many days did your back problem keep you from going to work (job, school, housework)?”
Core index	5	0–10	

^a Scales inverted (such that 1 = “very good/satisfied” and 5 = “very bad/dissatisfied”) before entered into the composite index score

- Fear-avoidance beliefs [26, 31]
- Depression (modified ZUNG self-rating depression scale [6, 14, 35])
- The “core-set” items [7, 16]

All the questionnaires had been cross-culturally adapted for the German language in previous studies [16, 18, 26].

From the core-set items (pain, function, symptom-specific well-being, general well-being, disability), a composite index score was constructed, as described previously [16]. Briefly, all scales were first linearly transformed into a 0–10 format. Pain intensity was already measured in this format, whilst function, and symptom-specific and general well-being were measured with a 1–5 point Likert scale (transformed according to the formula: category score marked by the patient -1×2.5). Disability (work) and disability (social role) were measured in days of work incapacity and restricted activity respectively over the last month

and could theoretically range from 0 to 31. These were recoded into five categories, to provide a similar 1–5 point scale as for the other items: (1) 0 days, (2) 1–7 days (3) 8–14 days (4) 15–21 days (5) >22 days.

The transformed core-item scores were averaged to form an unweighted composite core index that ranged from 0 to 10.

Data analysis and statistics

Descriptive data on the domain single items and the composite index score, as well as their test–retest reliability, validity and sensitivity to change, were published recently [16] and are therefore not reported here.

In order to predict the outcome after surgical treatment, each index item and the composite score, in turn, served as the dependent variable in a multivariate longitudinal regression analysis. In the longitudinal

hierarchical regression models, the baseline equivalent of the dependent variable was the first variable entered into the model (e.g. if pain was being predicted as the outcome measure, pain at baseline was the first variable entered into the analysis). This is the standard way of analysing longitudinal data and allows the prediction of change over and above that predicted by the association between the outcome variable and its value at baseline [12]. In the second step, the demographic variables, age and gender, were entered into the model. In the third step, the pain/medical variables were entered (duration of pain, number of previous spinal operations, number of vertebral levels treated, surgical procedure [dummy-coded variables]). We decided to include surgical procedure rather than diagnosis as a possible predictor variable in this step: these two variables are highly interrelated (and hence only one was to be entered), and we opted for surgical procedure, as it is potentially more relevant to the treatment outcome to consider what has actually been done than to consider diagnostic labels. Finally, in the fourth step, the psychosocial variables fear-avoidance beliefs (FABQ) about physical activity and about work, and depression (ZUNG) were entered into the model (for a similar multivariate regression model for predicting disability and work absence, see [31]).

Statistical significance was accepted at the $P < 0.05$ level.

Results

Participants/study flow

All patients who fulfilled the admission criteria ($n = 427$) were invited to participate; 256 (60%) patients agreed, and completed the first questionnaire 1–2 weeks before surgery. Many patients declared that they did not want to be bothered by such things as filling out long questionnaires before an impending operation; for many “it was simply too much”. Examination of the demographics of participants versus non-participants (for the Foundation Hospital only, in which access to such information was readily available) revealed that the group of patients that participated contained a slightly lower proportion of women (but not significantly; $P = 0.25$) and was somewhat older (62 vs 59 years; $P = 0.04$) compared with the group of patients that declined participation. The distribution of diagnoses did not differ significantly between the groups, but there was a slight tendency for discopathy and instability patients to be less well represented in the group under study.

Of the 256 patients who returned the baseline questionnaire, 25 patients being treated for a herniated disc were later excluded from analysis for a combination of reasons: they were somewhat more acute in their symptoms; they were recruited from one hospital only and relatively few in number (see [Methods](#)); and recent reports in the literature [5, 15] gave reason to believe they may differ in relation to their (lesser) propensity for non-medical factors to influence outcome (although being so few in number, any attempt to confirm such a bias by means of an independent analysis of this sub-group was untenable).

The data from a further 20 patients who had completed the baseline questionnaire were also discarded for various reasons: no follow-up questionnaire was sent to 7 patients due to administrative errors in the recall system, 3 patients due to uncertainties regarding the validity of the pre-op questionnaire (completion date/language-understanding), and 3 patients due to their involvement in a bigger research project requiring the completion of a similarly long questionnaire; 3 patients did not go on to receive the foreseen spine operation (cardiovascular comorbidity in 2, hip problem in 1); 2 patients developed other confounding medical problems (1 brain tumour, 1 hiatus hernia) and 2 patients underwent further spine operations, which precluded meaningful completion of the follow-up questionnaire 6 months after the index procedure.

Of the remaining 211 patients, 48 patients failed to respond to the 6-month follow-up questionnaire or returned a blank or largely incomplete questionnaire, whilst 163 patients returned a completed questionnaire. This gave a corrected response rate at follow-up of 77.3% (163/211). The demographic characteristics of the patients in the final study group are shown in [Table 2](#).

The data of the 48 dropouts were compared with those of the 163 responders with respect to age, gender, and all the variables that were predictors or outcome variables in the linear regression analyses. The dropout group was significantly younger than the group of responders [55.9 (SD 14.7) vs 61.4 (SD 14.3); $P = 0.02$] but the ratio of males to females was similar in the two groups [20:28 (41.7% males) and 69:94 (42.3% males), respectively; $P = 0.94$]. The dropout group showed significantly ($P = 0.023$) lower values for just one item of the Core index, general well-being (“How would you rate your quality of life?”). Otherwise, no differences between the groups reached statistical significance, although trends existed for the responders to have slightly higher baseline ZUNG depression and FABQ scores.

Table 2 Sample characteristics

	Total number (<i>N</i> = 163)	Subsample spinal stenosis (<i>N</i> = 97)
Sex (male/female)	69/94	47/50
Age (mean ± SD [range])	61.4 ± 14.3 (22.0–86.9)	66.8 ± 11.8 (30.0–86.9)
Diagnosis (number of patients)		
Spinal stenosis	97	97
Discopathy	39	–
Facet syndrome	5	–
Segmental instability	22	–
Surgical treatment (number of patients)		
Decompression	75	65
Decompression and fusion	53	31
Fusion	29	1
Other	6	–
Duration of back pain (number of patients)		
Missing values	4	4
Less than 1 week	0	0
1–4 weeks	0	0
1–3 months	5	5
3–12 months	20	12
1–2 years	15	8
2–5 years	51	28
More than 5 years	68	40
Number of previous spine operations (number of patients)		
Missing	2	1
None	105	71
1	33	16
2–3	17	8
>3	6	1
Number of levels treated (number of patients)		
1	81	32
2	72	40
3	28	22
4	4	3
5	0	0
6	2	0
Fear-Avoidance-Beliefs-Questionnaire (mean ± SD)		
FABQ1—work beliefs	24.3 ± 12.6	22.0 ± 12.8
FABQ2—activity beliefs	17.7 ± 5.6	17.1 ± 5.6
Zung Depression Scale (mean ± SD)	18.7 ± 9.3	17.1 ± 8.4

Since 97/163 of the study group were patients with spinal stenosis, a separate analysis was conducted for this group alone; the demographic details for this subgroup are shown in Table 2.

Changes in core item scores before and 6 months after surgery

The mean core-measures composite index score (score range 0–10) reduced from 7.0 (SD 1.6) before surgery to 5.0 (SD 2.4) after surgery ($P < 0.001$). All individual items showed highly significant improvements ($P < 0.001$).

Prediction of the core-measures composite index score

In predicting the core-measures composite score 6 months after surgery, the baseline (pre-operative) index score was a significant predictor when it was the only predictor variable in step 1 ($\beta = 0.447$, $P < 0.001$; Table 3). In step 2, age and gender were entered into the regression and showed no significant partial regression coefficients, i.e. did not contribute significantly to the predictive model. Also in step 3, when medical predictor variables entered the model, no individual predictor variable showed a significant

Table 3 Results of the hierarchical multiple regression analysis of the core measures composite index on demographic, medical, and psychosocial predictor variables

DV: core measures index at follow-up model	Unstandardized regression coefficients		Standardized regression coefficients		T	P	95% CI for B		Correlations			Explained variance in DV		
	B	SE	β	SE			CI low	CI high	Zero-order	Partial	Part		R^2	Change in R^2
Step 1														
Baseline scores	0.146	0.933	0.447	0.083	0.156	0.876	-1.702	1.994						
of DV	0.703	0.131			5.378	0.000	0.444	0.962	0.447	0.447	0.447	0.200		
Step 2														
Demographics added	1.567	1.491	0.446	0.083	1.051	0.296	-1.387	4.521						
Constant	0.702	0.131	-0.082	0.085	5.358	0.000	0.442	0.961	0.447	0.449	0.446			
Core measures index pre-op	-0.014	0.014	-0.078	0.085	-0.967	0.336	-0.042	0.014	-0.072	-0.090	-0.080			
Age	-0.389	0.421			0.357	0.357	-1.223	0.444	-0.061	-0.086	-0.077	0.010		
Gender														
Step 3														
Medical variables added	-0.112	1.908	0.413	0.086	-0.059	0.953	-3.895	3.670						
Constant	0.650	0.135	-0.031	0.093	4.807	0.000	0.382	0.918	0.447	0.420	0.399			
Core measures index pre-op	-0.005	0.016	-0.087	0.088	-0.330	0.742	-0.036	0.026	-0.072	-0.032	-0.027			
Age	-0.429	0.437			0.328	0.328	-1.296	0.437	-0.061	-0.094	-0.082			
Gender	0.200	0.198	0.089	0.088	1.012	0.314	-0.192	0.592	0.115	0.097	0.084			
Duration of pain pre-op	0.281	0.276	0.094	0.092	1.019	0.310	-0.266	0.828	0.216	0.098	0.085			
No. of previous spine ops	-0.151	0.231	-0.059	0.090	-0.653	0.515	-0.609	0.308	-0.010	-0.063	-0.054			
No. of levels treated	0.229	0.625	0.037	0.101	0.366	0.715	-1.009	1.467	0.069	0.035	0.030			
Treatment: only fusion	1.805	1.102	0.148	0.090	1.638	0.104	-0.379	3.990	0.217	0.156	0.136			
Treatment: other	0.324	0.488	0.063	0.094	0.665	0.507	-0.642	1.291	-0.039	0.064	0.055	0.045		
Treatment: decompression and fusion														
Step 4														
Psychosocial variables added	0.248	1.840	0.190	0.107	0.135	0.893	-3.399	3.896						
Constant	0.299	0.168	0.010	0.090	1.786	0.077	-0.033	0.632	0.447	0.172	0.142			
Core measures index pre-op	0.002	0.015	-0.118	0.085	0.108	0.914	-0.028	0.032	-0.072	0.011	0.009			
Age	-0.588	0.422	0.062	0.085	-1.393	0.166	-1.424	0.249	-0.061	-0.135	-0.111			
Gender	0.140	0.191	0.116	0.089	0.733	0.465	-0.239	0.518	0.115	0.071	0.058			
Duration of pain pre-op	0.348	0.266	0.002	0.087	1.309	0.194	-0.179	0.875	0.216	0.127	0.104			
No. of previous spine ops	-0.151	0.222	-0.059	0.087	-0.681	0.497	-0.591	0.289	-0.010	-0.066	-0.054			
No. of levels treated	0.012	0.605	0.002	0.097	0.021	0.984	-1.187	1.212	0.069	0.002	0.002			
Treatment: only fusion	1.296	1.064	0.106	0.087	1.218	0.226	-0.814	3.405	0.217	0.118	0.097			
Treatment: other	0.309	0.466	0.060	0.090	0.662	0.510	-0.616	1.233	-0.039	0.064	0.053			
Treatment: decompression and fusion	0.041	0.020	0.206	0.098	2.113	0.037	0.003	0.080	0.401	0.202	0.168			
FABQ work baseline	0.000	0.043	0.000	0.096	0.004	0.996	-0.085	0.086	0.280	0.000	0.000			
FABQ activity baseline	0.069	0.027	0.257	0.102	2.528	0.013	0.015	0.122	0.431	0.240	0.201	0.339		
ZUNG depression baseline												0.084**		

DV dependent variable (i.e. the outcome being predicted)

^a The index comprised five items: pain symptoms, back function, general well-being, symptom-specific well-being, and disability; it was calculated as the mean of the five item scores after each been transformed to a ten-point (0–10) scale

*** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$

partial regression coefficient; taken together as a block of predictors, the duration of pain, the number of previous surgeries, the number of treated levels, and the kind of treatment added 4.5% of variance explanation (ns) to the model. In step 4 of the hierarchical regression analysis, activity-related and work-related fear-avoidance beliefs and depression, as measured with the Zung scale, entered the model. FABQ work and ZUNG each showed significant beta coefficients (FABQ work: 0.206, $P = 0.037$; ZUNG: 0.257, $P = 0.013$). In the final model, the unique contribution of the block of medical and psychosocial variables was calculated by multiplying the standardized partial regression coefficients by the zero-order correlation between each predictor variable and the index, and summing these afterwards within each block. The unique contribution of medical predictors was 5.4%, whereas the psychosocial variables explained nearly 20% of the variance in the index (Table 4). Similar results were found when just the spinal stenosis patients were analyzed (2.3% medical predictor variables, 18.3% psychosocial predictor variables).

Table 4 also shows the results for prediction of the index when the single items of the index were z -standardized and the index was calculated as the mean score of z -standardized item values. Z -standardisation involves converting individual scores into measurements of standard deviations above or below the

sample mean (which after transformation then becomes zero, with a SD of 1), making variables with different units or scales of measurement more directly comparable. Within the present study, it would serve to level out possible differences in the response range of the single items and might therefore be seen as a psychometrically better means of analysis. As shown in Table 4 (last column), regression of the z -standardized item-index on predictor variables showed comparable results to those for the non-standardized item-index, although the unique variance explained by psychosocial predictor variables was somewhat higher (25.7% compared with 19.4% for the combined predictor model).

Altogether, the combination of baseline symptoms, medical variables, and psychosocial factors explained 34% variance in the core-index score (regardless of whether raw or z -standardized scores were used) (Table 4, combined model).

For better comparison of the sets of “medical” and “psychosocial” predictor variables, three separate regression models were constructed (Table 4). The “medical” and “psychosocial” predictor models each included just the corresponding set of predictors, controlling for demographics and the baseline value of the outcome variable in question. Each set shows its maximum predictive power in these stand-alone models. In the third model, both sets of predictor variables were entered into the model simultaneously,

Table 4 Unique variance explained in the final regression models by sets of medical predictors, psychosocial predictors, and combined medical and psychosocial predictors

	Predictor variables included in the model	Pain symptoms	Back function	Symptom specific well-being	General well-being	Disability	Core measure index	Core measure index z -scored
“Medical” predictor model	DV at baseline	0.129	0.049	0.040	0.151	0.165	0.185	0.135
	Age, gender	0.001	0.006	0.001	0.007	0.013	0.008	0.006
	Medical predictors	0.073	0.047	0.085	0.018	0.048	0.066	0.055
	Total	0.203	0.102	0.126	0.176	0.226	0.259	0.196
“Psychosocial” predictor model	DV at baseline	0.093	0.012	0.030	0.070	0.095	0.093	0.042
	Age, gender	0.002	0.008	0.005	0.006	0.014	0.009	0.007
	Psychosocial predictors	0.138	0.202	0.150	0.166	0.213	0.202	0.267
	Total	0.233	0.222	0.185	0.242	0.322	0.304	0.316
Combined medical and psychosocial predictor model	DV at baseline	0.094	0.010	0.029	0.062	0.104	0.085	0.041
	Age, gender	0.001	0.007	0.000	0.008	0.010	0.007	0.004
	Medical predictors	0.064	0.036	0.070	0.018	0.024	0.054	0.055
	Psychosocial predictors	0.128	0.199	0.140	0.171	0.203	0.194	0.257
	Total	0.287	0.252	0.239	0.259	0.341	0.340	0.340
Same model only stenosis ($N = 97$)	DV at baseline	0.198	0.019	0.047	0.081	0.158	0.085	0.072
	Age, gender	0.003	0.006	0.014	0.003	0.012	0.002	0.000
	Medical predictors	0.059	0.015	0.035	0.047	0.033	0.023	0.018
	Psychosocial predictors	0.144	0.151	0.097	0.163	0.174	0.183	0.215
	Total	0.404	0.191	0.193	0.294	0.377	0.289	0.305

and they hence had to ‘compete’. In the combined model, the variance explained by each set of variables is unique, in that it is controlled for the other variables: if this unique contribution were much lower than in the corresponding stand-alone model, it would indicate that the medical and psychosocial predictor sets were not complementary in their prediction but rather were predicting very much the same inter-individual differences in outcome, i.e. delivering overlapping information. Notably, this was not the case. The variance explained by the sets in the stand-alone models was almost identical to their unique contribution in the combined model (Table 4).

Prediction of the individual domain items

Hierarchical multiple regression analyses showed that the medical variables were better in predicting pain (6.4% variance explained) and symptom-specific well-being (7.0%) than in predicting back function (3.6%), general well-being (1.8%) or disability (2.4%). The inverse pattern emerged for the psychosocial predictors. The latter explained a higher proportion of variance in back function (19.9%), general well-being (17.1%), and disability (20.3%) than pain (12.8%) or symptom-specific well-being (14.0%). This pattern did not change much when “medical” and “psychosocial” predictor sets were analysed separately (Table 4).

Discussion

The present study confirmed that the outcome after spinal surgery, as measured using the recently validated core-measures questionnaire [7, 16, 22], can be predicted by some of the variables most commonly considered in the literature to represent determinants of outcome [15]. The questionnaire has recently been incorporated into the Spine Society of Europe’s “Spine Tango Spine Surgery Registry” (<http://www.spinetan-go.com>) and is being recommended for use on a prospective basis for all patients undergoing spinal surgery in the institutions contributing to the registry. A knowledge of the factors that might influence outcome, as assessed with this instrument, is hence of great interest.

Together with the baseline symptoms, the medical and psychosocial predictor variables examined in the present study were able to explain a considerable proportion of the variance in outcome as assessed by the composite-index score—34%. In terms of effect sizes, this would be considered large (i.e. an effect size

of 0.51, where the effect size $f^2 = R^2/1 - R^2$, and f^2 values of 0.02, 0.15, and 0.35 are considered to indicate small, medium, and large effects, respectively [3]). Even when considering only the medical (5.4%) and psychosocial predictor sets (19.4%, see Table 4), these together explained 24.8% of the variation in the index, which in terms of an effect size (0.33) is also large and comparable with the findings of high quality prospective studies reported in the recent literature (see review in [15]). With the limitations of any biologic measurement, it is often possible to identify only 25–50% of the variance of a relationship [30]. Thus, the composite index proved itself to be a responsive instrument with respect to risk factor-related change. This complements previous findings in relation to the instrument’s responsiveness (sensitivity to change) after treatment, where effects sizes (0.95–1.30) exceeding those of many longer outcome questionnaires were found [16, 23, unpublished findings].

Various factors representing the different domains (medical and psychosocial) were identified as unique, independent predictors of outcome. Age and gender did not make any significant contribution to the explanatory models for any of the outcome items or the composite score. Whilst these variables have been identified as significant predictors of outcome in some retrospective studies (in which a limited number of “readily available” risk factors are typically examined), the majority of high quality prospective studies do not support a predictive role for them [15]. One of the risk factors commonly identified in previous studies, the duration of symptoms [11, 19, 21, 24, 29, 32, 33], was not shown to be a significant predictor in the present study. This may have been due to the paucity of acute patients in the study sample that would otherwise have provided a wider spectrum of values for this independent variable; a very high proportion of the whole sample (94%) were already “chronic” pain sufferers, according to the standard definition of >3 months pain.

Many work-related factors, such as worker’s compensation, disability claims, work status, and the duration of sick leave, have been identified as predictors of surgical outcome (see [15]). These factors were not examined in detail in the present study, but we recommend their inclusion in future risk-factor studies involving the core measures, particular in relation to the item “disability” (assessed by the number of cut-down days/time off work in the preceding month). In the present study, in the prediction of disability, 30 individuals had missing data, because they did not work at baseline and/or follow-up. In the remaining occupational group, fear-avoidance beliefs about work

was the most significant predictor of “disability”, confirming the findings of previous studies in which more elaborate disability scales were used [17, 26, 31].

In the present study, depression was a consistent predictor of many of the individual outcome items and also of the composite core-index measure. Over the last two decades, the role of depression/psychological distress as a predictor of surgical outcome has been the subject of much study and discussion. Currently, the only consensus appears to be that there is no consensus with regard to the findings, with as many studies as not, finding an association between depression and outcome (see [15]). Some suggest that the poor results of surgery reported in psychologically disturbed patients may reflect intervention in patients who did not have surgically remediable pathology: distress may increase the pressure for surgery, and inappropriate symptoms and signs may obscure the physical assessment, leading to a mistaken diagnosis of a surgically treatable lesion [30]. In this sense, the uncertain indication rather than psychological distress, per se, would be the factor responsible for the poor outcome, and would no doubt further exacerbate the distress of the patient after surgery [28], hence explaining the commonly reported association between outcome and post-operative psychological status (see [15]). To an extent, this theory has been verified by the recent studies of Carragee et al. (see [4]). This group showed that patients with acute and subacute sciatica in association with a clearly identifiable, severe disc herniation had a very high chance of dramatic and lasting improvement with surgery, and that standard psychological screening tests failed to predict outcome in these patients. Even severe emotional distress in patients who underwent appropriate early surgical intervention did not correlate with adverse outcomes, although the same psychometric profile in patients with chronic sciatic pain and disability did predict worse outcomes. These authors concluded that, with prolonged pain and emotional distress, adverse and possibly self-perpetuating psychological and social changes may significantly decrease the impact of surgery [4]. In the present study, more than 80% of the patients had had pain for more than 1 year, which most likely explains the significance of depression in predicting surgical outcome in this group.

The multiple predictors identified for the composite index show that: (1) the index is sensitive to the multidimensional risk factors that are most consistently described in the outcomes literature [15], and (2) the theoretical notion of building an index out of multiple relevant outcome dimensions, each of which had its own specific risk-factor profile in single item

analyses, was successful in maintaining those same specific aspects within the combined index. In other words, the multi-dimensionality was reflected in a risk factor profile that constituted a combination of the profiles from the single item analyses. The potential danger in making an aggregate index is that it might equalize out the specific characteristics of the individual dimensions. However, this was clearly not the case here, and the composite index represented its individual facets very well. Nonetheless, we consider it prudent to recommend that, in future studies, the treatment effects for the individual items be examined in addition to those of the composite core measures index.

Certain limitations of the present study are worthy of mention. Only 60% of the patients who were invited to participate actually volunteered to do so. This potentially threatens the validity of the study and the ability to extrapolate the findings to the “typical” spine surgery patient. Nonetheless, we believe that those who did participate represented reasonably well the individuals seeking care in our institutions, since there were no relevant differences in their demographic characteristics compared with the non-responders, and their baseline and follow-up core index scores were comparable to those of patients with the same diagnoses that are now assessed with the core measures on a routine basis in one of the participating hospitals [unpublished findings]. Of the group that did participate, those that subsequently dropped out at follow-up were significantly younger, and rated their quality of life at baseline as significantly worse, than those who returned their follow-up questionnaire; they also had a non-significant tendency to be more depressed and fearful. However, since non-responders at follow-up are generally considered to include a higher proportion of patients with a poor outcome [13, 20], rather than challenging the validity of our conclusions this would serve instead to confirm our hypothesis that a less healthy psychological profile is associated with a poorer outcome.

In conclusion, further to our previous studies establishing the reliability, validity, and sensitivity to change of the core-index [16], we have now shown that the variance in its value after surgery could be significantly explained by the most commonly reported medical and psychosocial predictor variables. Moreover, the index was represented well by its individual items, as documented by the fact that certain predictor variables were more specific for some domains, such as the psychosocial predictors for “disability” and “function” and “general well-being”, rather than for “pain” and “symptom-specific well-being”. We believe

that this further substantiates the recommendation for the widespread and consistent use of the core measures index in clinical trials, multicentre studies, routine quality management, and surgical registry systems.

References

- Block AR, Ohnmeiss DD, Guyer RD, Rashbaum RF, Hochschuler SH (2001) The use of presurgical psychological screening to predict the outcome of spine surgery. *Spine J* 1(4):274–282
- Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders: introduction. *Spine* 25(24):3097–3099
- Bortz J, Doering N (2002) *Forschungsmethoden und evaluation*, 3rd edn. Springer, Berlin Heidelberg New York
- Carragee EJ (2001) Psychological screening in the surgical treatment of lumbar disc herniation. *Clin J Pain* 17(3):215–219
- Carragee EJ, Han MY, Suen PW, Kim D (2003) Clinical outcomes after lumbar discectomy for sciatica: the effects of fragment type and anular competence. *J Bone Joint Surg Am* 85-A(1):102–108
- CIPS - Collegium Internationale Psychiatricae Salarum (Hrsg) (1986) *Internationale Skalen für die Psychiatrie*, 3rd edn. Beltz, Weinheim
- Deyo RA, Battie M, Beurskens AJHM, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G (1998) Outcome measures for low back pain research. A proposal for standardized use. *Spine* 23(18):2003–2013
- Dionne CE, Von Korff M, Koepsell TD, Deyo RA, Barlow WE, Checkoway H (1999) A comparison of pain, functional limitations, and work status indices as outcome measures in back pain research. *Spine* 24(22):2339–2345
- Hagg O, Fritzell P, Ekselius L, Nordwall A (2003) Predictors of outcome in fusion surgery for chronic low back pain. A report from the Swedish Lumbar Spine Study. *Eur Spine J* 12(1):22–33
- Howe J, Frymoyer JW (1985) The effects of questionnaire design on the determination of end results in lumbar spinal surgery. *Spine* 10(9):804–805
- Junge A, Dvorak J, Ahrens S (1995) Predictors of bad and good outcomes of lumbar disc surgery. A prospective clinical study with recommendations for screening to avoid bad outcomes. *Spine* 20(4):460–468
- Kasl SV, Jones BA (2003) An epidemiological perspective on research design, measurement, and surveillance strategies. In: *Handbook of occupational health psychology*. APA, Washington, pp 379–398
- Kim J, Lonner JH, Nelson CL, Lotke PA (2004) Response bias: effect on outcomes evaluation by mail surveys after total knee arthroplasty. *J Bone Joint Surg Am* 86-A(1):15–21
- Main CJ, Waddell G (1984) The detection of psychological abnormality in chronic low back pain using four simple scales. *Curr Concepts Pain* 2:10–15
- Mannion AF, Elfering A (2006) Predictors of surgical outcome and their assessment. *Eur Spine J* 15(Suppl 1):S93–S108
- Mannion AF, Elfering A, Staerkle R, Junge A, Grob D, Semmer NK, Jacobshagen N, Dvorak J, Boos N (2005) Outcome assessment in low back pain: how low can you go? *Eur Spine J* 14(10):1014–1026
- Mannion AF, Junge A, Taimela S, Muntener M, Lorenzo K, Dvorak J (2001) Active therapy for chronic low back pain: part 3. Factors influencing self-rated disability and its change following therapy. *Spine* 26(8):920–929
- Mannion AF, Muntener M, Taimela S, Dvorak J (1999) A randomised clinical trial of three active therapies for chronic low back pain. *Spine* 24(23):2435–2448
- Ng LC, Sell P (2004) Predictive value of the duration of sciatica for lumbar discectomy. A prospective cohort study. *J Bone Joint Surg Br* 86(4):546–549
- Norquist BM, Goldberg BA, Matsen FA 3rd (2000) Challenges in evaluating patients lost to follow-up in clinical studies of rotator cuff tears. *J Bone Joint Surg Am* 82(6):838–842
- Nygaard OP, Kloster R, Solberg T (2000) Duration of leg pain as a predictor of outcome after surgery for lumbar disc herniation: a prospective cohort study with 1-year follow up. *J Neurosurg Spine* 92(2):131–134
- Pellise F, Alvarez L, Escudero O, Pont A, Ferrer M (2003) Metric characteristics of the six-question “core set” in the evaluation of back pain. *Eur Spine J* 12(Suppl 1):S12–S13
- Pellise F, Ferrer M, Escudero O, Alvarez L, Pont A, Alonso J, Deyo R (2005) Reliability and responsiveness of a “minimum core-set” in the evaluation of patients with chronic low back pain. *Eur Spine J* 14(Suppl 1):S18
- Rothoerl RD, Woertgen C, Brawanski A (2002) When should conservative treatment for lumbar disc herniation be ceased and surgery considered? *Neurosurg Rev* 25(3):162–165
- Schade V, Semmer N, Main CJ, Hora J, Boos N (1999) The impact of clinical, morphological, psychosocial and work-related factors on the outcome of lumbar discectomy. *Pain* 80(1–2):239–249
- Staerkle R, Mannion AF, Elfering A, Junge A, Semmer NK, Jacobshagen N, Grob D, Dvorak J, Boos N (2004) Longitudinal validation of the fear-avoidance beliefs questionnaire (FABQ) in a Swiss–German sample of low back pain patients. *Eur Spine J* 13(4):332–340
- Stromqvist B (2002) Evidence-based lumbar spine surgery. The role of national registration. *Acta Orthop Scand Suppl* 73(305):34–39
- Tandon V, Campbell F, Ross ER (1999) Posterior lumbar interbody fusion. Association between disability and psychological disturbance in noncompensation patients. *Spine* 24(17):1833–1838
- Trief PM, Grant W, Fredrickson B (2000) A prospective study of psychological predictors of lumbar surgery outcome. *Spine* 25(20):2616–2621
- Waddell G, Morris EW, Di Paola MP, Bircher M, Finlayson D (1986) A concept of illness tested as an improved basis for surgical decisions in low-back disorders. *Spine* 11(7):712–719
- Waddell G, Newton M, Henderson I, Somerville D, Main CJ (1993) A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 52(2):157–168
- Woertgen C, Holzschuh M, Rothoerl RD, Brawanski A (1997) Does the choice of outcome scale influence prognostic factors for lumbar disc surgery? A prospective, consecutive study of 121 patients. *Eur Spine J* 6(3):173–180
- Woertgen C, Rothoerl RD, Breime K, Altmeyen J, Holzschuh M, Brawanski A (1999) Variability of outcome after lumbar disc surgery. *Spine* 24(8):807–811
- Zanoli G, Stromqvist B, Padua R, Romanini E (2000) Lessons learned searching for a HRQoL instrument to assess the results of treatment in persons with lumbar disorders. *Spine* 25(24):3178–3185
- Zung WW (1965) A self-rating depression scale. *Arch Gen Psychiatry* 12:63–70