

SF-36 scores in degenerative lumbar spine disorders

Analysis of prospective data from 451 patients

Gustavo Zanolli^{1,2}, Bo Jönsson¹ and Björn Strömqvist¹

Departments of Orthopedics, ¹Lund University Hospital, Lund, Sweden, ²University of Ferrara, Italy
Correspondence GZ: gzanoli@globeweb.org
Submitted 05-02-05. Accepted 05-10-09

Background When using Health-Related Quality of Life (HRQoL) in assessing outcomes of treatment, normative data for different diagnoses are needed to allow cross-comparisons of existing and future studies. We determined the SF-36 scores in patients with surgical lumbar spine problems.

Methods This is a prospective observational study of consecutive surgical patients in one institution. In addition to SF-36 questionnaire responses, local pain, radiating pain, analgesic intake and walking ability were recorded, together with several other demographic variables. 451 patients (50% women) with median age 52 (13–88) years, operated from 1998 through 2002, were included in the study.

Results Preoperative SF-36 scores were significantly lower than those derived from previously published material (the general population, nonspecific low back pain (LBP) patients, other samples of non-LBP patients), also with the use of norm-based scoring. Sick-leave and worker's compensation seemed to affect perceived Health-Related Quality of Life (HRQoL), but smoking habits did not. Some SF-36 domains showed a possible discriminating pattern between diagnoses.

Interpretation HRQoL reported by patients scheduled for lumbar spine surgery was much worse than for the normal population and for LBP patients. The normative SF-36 values provided may be used as a benchmark in future studies.

in assessing outcomes of spinal surgery (Bombardier 2000a). The initial literature on the subject was largely retrospective (Padua et al. 2001), but nowadays HRQoL questionnaires are included in most of the prospective clinical trials published. This is considered to be particularly useful, not only because it can increase the clinical significance of a study but also because it can allow comparisons across studies, if standard questionnaires (i.e. validated and internationally accepted questionnaires) are used (Zanolli et al. 2000).

One of the most frequently used questionnaires for HRQoL evaluation in spinal pathology is SF-36 (Ware and Sherbourne 1992). To confirm this, we performed a search for articles which deal with "SF-36" and "spine" in MEDLINE and restricted it to publication year 2004; 35 references were retrieved. The reasons for this degree of popularity have been highlighted and discussed elsewhere (Bombardier 2000b, Ware et al. 2000): the SF-36 questionnaire appears to achieve the best balance between length, reliability, validity, responsiveness, and experience in large populations of patients with back pain.

The SF-36 questionnaire is a multi-purpose, short-form health survey with 36 questions. It yields an 8-scale profile of functional health and well-being scores. The 8 health concepts were selected from 40 included in the Medical Outcomes Study (MOS) (Stewart and Ware 1992). They represent Physical Function (PF), Role Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Function (SF), Role Emotional (RE), and Mental Health (MH).

There is an increasing consensus in the use of Health-Related Quality of Life (HRQoL) measures

Two of the most interesting features of SF-36 are that it has been validated in many different languages and countries, and that national normative data are available. These data usually include normal values for healthy persons, as well as for a series of common pathological conditions such as nonspecific low back pain (LBP). For example, Swedish normative data were obtained as a result of the international IQOLA project out of a nationwide sample of 8,930 individuals (Sullivan et al. 1994).

Clinical studies often report baseline SF-36 scores for spine disorders which are much lower than the published normal values when either healthy individuals or LBP patients have been sampled (Dahl et al. 2001, Coulter et al. 2002). One possible explanation for this is that patients treated and included in specialty trials are usually more severely affected than a generic sample of individuals who report having low back pain on a questionnaire. Thus, the nonspecific LBP sample population presented in the SF-36 manual (Ware et al. 1993) is probably not always representative of—or at least not sufficient for comparison with—surgical cases, for example.

On the other hand, data from these published studies is usually too specific or not large enough to be used as a source of potentially “normal” values for any given spinal condition. To allow cross-comparisons of existing retrospective studies and for a better interpretation of baseline characteristics of patients in prospective studies, normative data for different specific diagnoses are needed.

We determined the preoperative SF-36 scores in a large sample of patients who were operated on for lumbar spine problems. Data were extracted from a large prospective, consecutive database of patients operated on for low back problems at our institution, and they were analyzed in detail for the 5 most frequent surgical diagnoses.

Patients and methods

Data were obtained using a prospective observational registration protocol (Strömqvist and Jönsson 1993) of patients operated at our institution and included in the Swedish National Lumbar Spine Surgery Registry (Strömqvist et al. 2001). Between 1998 and 2002, 477 patients undergoing

elective lumbar surgery at the spine section of the Department of Orthopedics, Lund University Hospital, were included in the protocol for the Swedish Lumbar Spine Registry which (as of 1998) contains the Swedish version of the SF-36 questionnaire. 451/477 patients (95%) returned the forms, and the data are presented here.

Initial information at baseline (the day before surgery) included age, sex, and duration of preoperative back and leg pain. The use of several diagnostic techniques (CT, MRI, myelography, myeloCT, nerve root injections) during the diagnostic work-up was recorded as a binary variable (yes/no). Questionnaires were also used to collect information on working situation, smoking habits, walking ability, and analgesic intake, and to record visual analog scale (VAS) scores for back pain and for radiating pain.

Statistics

Parametric tests (mean, SD and 95%CI) were used when only SF-36 variables were involved in the analysis. When comparison with other variables was done, for which no assumption of normal distribution could be made, non-parametric tests (Mann-Whitney, Spearman rho) were used. Statistical analysis was performed using SPSS for Windows version 11.0. 95% confidence intervals for median values were calculated according to Bland (Conover 1980).

Results

All baseline characteristics of patients included in the study are shown in Table 1.

A preliminary check was performed (Mann-Whitney test, with significance level set at $p < 0.05$), to ensure that the selection of patients (responders and non-responders) had produced no bias regarding demographics and preoperative variables. The 5 most frequent diagnoses were: herniated nucleus pulposus (HD, 31%), central stenosis (CS, 27%), lateral stenosis (LS, 12%), isthmic spondylolysis (SO, 11%), and degenerative disk disease (DDD, 13%).

Surgical interventions ranged from discectomy (percutaneous 7.3%, microscopic 12%, or conventional 11%) to decompression (alone 41%, with

Table 1. Characteristics of study population. HD, lumbar disc herniation; CS, central stenosis; DDD, degenerative disk disease; SO, spondylolisthesis; LS, lateral stenosis; Other, miscellaneous and combined diagnoses

		Total (n = 477)	Responders (n = 451)	Non-responders (n = 26)	P-value Mann-Whitney
Sex	Male	240 (50%)	224 (50%)	16 (62%)	0.2
	Female	237 (49%)	227 (50%)	10 (39%)	
Age ^a		52 (50–55)	52 (50–54)	55 (47–71)	0.2
Duration (months) of back pain ^a		17 (15–19)	18 (15–19)	15 (7–53)	0.7
	leg pain ^a	15 (14–18)	15 (13–18)	17 (7–40)	0.6
Smoking	Yes	122 (26%)	119 (26%)	3 (12%)	0.2
	No	347 (73%)	329 (73%)	18 (69%)	
	Missing	8 (2%)	3 (1%)	5 (19%)	
Diagnosis	HD	150 (31%)	144 (32%)	6 (23%)	0.5
	CS	130 (27%)	121 (27%)	9 (35%)	
	DDD	63 (13%)	57 (13%)	6 (23%)	
	SO	54 (11%)	52 (12%)	2 (8%)	
	LS	57 (12%)	55 (12%)	2 (8%)	
	Other	23 (5%)	22 (5%)	1 (4%)	

^a Values are median (95% CI)

Table 2. Mean SF-36 scores for the 5 most frequent surgical diagnoses. PF, Physical Function; RP, Role Physical; BP, Bodily Pain; GH, General Health; VT, Vitality; SF, Social Function; RE, Role Emotional; MH, Mental Health. Mean score is the group mean

Diagnosis	n (%)	Age	Men (%)	PF	RP	BP	GH	VT	SF	RE	MH
Herniated disc	144 (32)	41 (39–45)	55								
Mean score				44	13	25	66	39	53	36	62
SD				20	27	16	23	22	27	43	22
95% CI				40–47	8–17	22–28	62–70	36–43	49–58	29–43	58–66
Central stenosis	121 (27)	70 (67–71)	42								
Mean score				30	8	27	59	40	57	33	63
SD				20	21	16	23	22	27	45	23
95% CI				26–33	4–11	24–29	54–63	36–45	53–62	25–41	59–68
Degen. disc disease	57 (13)	49 (45–56)	51								
Mean score				30	4	18	56	30	43	28	57
SD				16	14	12	26	21	27	42	20
95% CI				26–34	1–8	15–22	49–63	25–36	36–50	17–39	5–62
Spondylolisthesis	52 (12)	46 (41–53)	35								
Mean score				43	16	30	64	45	70	54	68
SD				20	32	18	25	25	28	45	21
95% CI				38–49	7–25	25–35	57–71	38–52	62–78	41–66	62–74
Lateral stenosis	55 (12)	53 (51–60)	64								
Mean score				39	11	29	57	37	57	36	61
SD				19	23	14	21	20	26	42	24
95% CI				34–44	5–17	26–33	51–62	32–43	50–64	24–47	55–68
Total sample	451 (100)	52 (50–55)	49.7								
Mean score				37	10	26	61	39	55	36	62
SD				21	24	16	24	22	28	44	23
95% CI				35–39	8–13	24–27	59–63	37–41	53–58	32–40	60–64

non-instrumented fusion 0.7%, or with instrumented fusion 12%) to fusion (anterior 2.9%, non-instrumented posterior 5.8%, or instrumented posterior 6.4%).

Mean SF-36 scores and standard deviations are reported separately for the 5 diagnostic categories (Table 2). Most of the domain values, and certainly the first 3 physical ones, are considerably lower

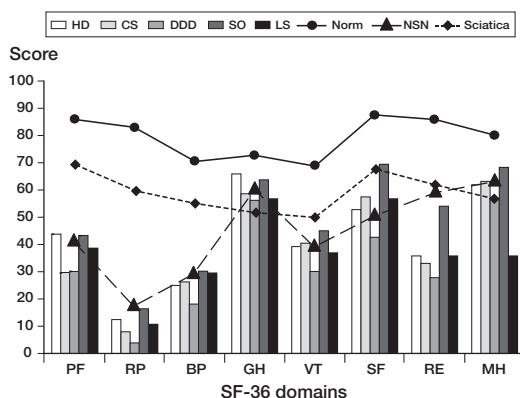


Figure 1. SF-36 profiles for the 5 diagnostic categories (bars) compared with normative values for the Swedish normal population (Sullivan et al. 1994) (unbroken lines), a sample of patients with sciatica (Ware et al. 1993) (dotted lines), and a survey of patients from the US National Spine Network (NSN) (Fanuele 2000) (broken lines).

SF-36 domains: PF, Physical Function; RP, Role Physical; BP, Bodily Pain; GH, General Health; VT, Vitality; SF, Social Function; RE, Role Emotional; MH, Mental Health.

Diagnoses: HD, lumbar disk herniation; CS, central stenosis; DDD, degenerative disk disease; SO, spondylolisthesis; LS, lateral stenosis.

than normative values for an age-matched sample and a sample of patients with non-specific sciatica as reported in the SF-36 manual (Ware et al. 1993, Sullivan et al. 1994) (Figure 1).

A graphical visualization of the SF profiles with mean differences is shown in Figures 1 and 2. Some SF-36 domains showed a possible discriminating pattern when analyzed separately for the five diagnoses (Figure 3). Among other preoperative variables studied, age showed a similar discriminating performance, whereas duration of back and leg pain was significantly shorter for HD alone (median 9 months) than for any other diagnosis (median 24 months) ($p < 0.01$).

SF-36 scores were tested regarding correlation to other preoperative clinical variables (walking ability, visual analog scale (VAS) scores for back pain and for radiating pain, analgesic intake). Most correlation coefficients were statistically significant, although Spearman's rho was not very high in general (less than 0.4 in all cases except the correlation between the Physical Function (PF) domain of SF-36 and the direct question on walking ability, where $\rho = 0.66$).

SF-36 profiles were compared between several subgroups, categorized for number of diagnostic

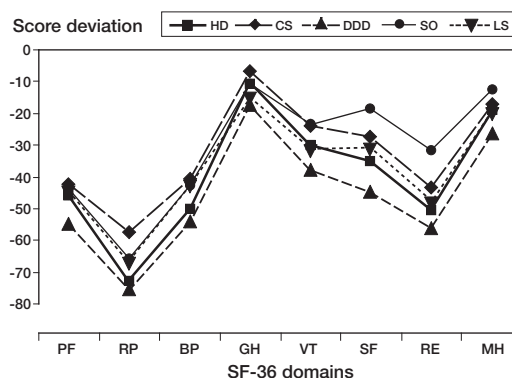


Figure 2. Mean deviations from the norms of the Swedish general population regarding the eight SF-36 domains for the five diagnostic categories (age- and sex-adjusted scoring). Abbreviations, see Figure 1.

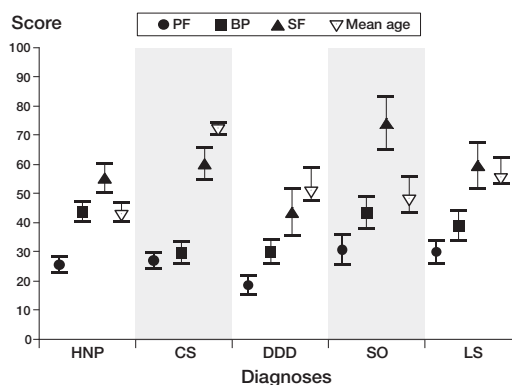


Figure 3. Mean values (with 95% CI) of the PF, BP, SF domains and median age (with 95% CI) for the 5 diagnostic categories. Abbreviations, see Figure 1.

tests performed before the operation (only one significant difference in the General Health (GH) domain, lower value for those who underwent more diagnostic tests) and for preoperative working situation (significantly lower values in the Role Physical (RP) and Bodily Pain (BP) domains for heavy workers, significantly higher value in the Role Emotional (RE) domain for light workers, with medium-heavy workers showing an intermediate pattern). More evident differences were found between the working and non-working group (significantly higher values in the PF, GH, and RE domains for patients who were working at the time of hospital admission). 8.6% of the patients were in long-term worker's compensation programs at the time of admission (WCP group), and 60% were

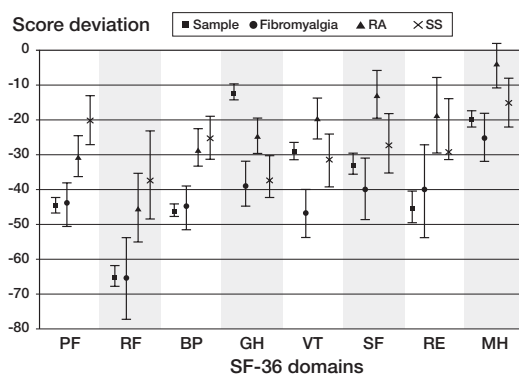


Figure 4. Mean deviations (with 95% CI) from the norms of the Swedish general population regarding the eight SF-36 domains (age- and sex-adjusted scoring) for the total sample of this study. This is compared with data from Strömbeck et al. (2000) for a group of women with Fibromyalgia ($n = 44$, mean age 53 years), rheumatoid arthritis ($n = 59$, mean age 55), and primary Sjögren syndrome ($n = 42$, mean age 59). Abbreviations, see Figure 1.

on sick leave before the operation (SL group). The WCP group had significantly lower values in the PF, RP and GH domains, whereas the SL group had significantly lower values in all domains except GH. Smoking habits had no apparent effect on pre-operative HRQoL in this group of patients, since the profiles of smokers and non-smokers could be superimposed.

For comparison, data from selected previously published studies with non-LBP diagnoses are presented graphically in Figures 4 and 5.

Discussion

This is mainly an observational report, and we see two main reasons for interest in our observations: (1) they provide data for comparison between existing and future studies with different population samples and diagnoses, and (2) they serve as a source of reference information for researchers planning clinical trials, for example when performing calculations of sample size.

A preliminary check was performed on the patients who had received the questionnaire but had not completed it, to exclude any possible selection bias. No significant difference was found in the preoperative variables (Table 1). Even though this is not a clinical study, knowing the percentage of complete data is important to allow extrapolation

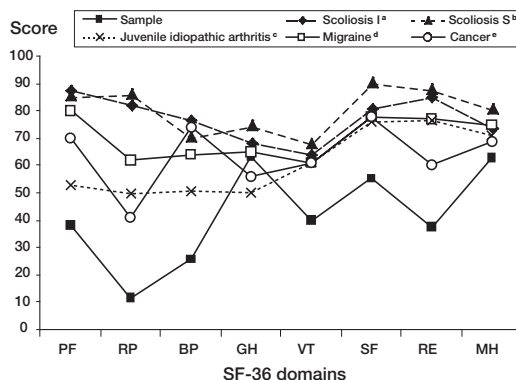


Figure 5. SF-36 profiles for the total sample of this study compared with data from long-term follow-up studies of patients operated for scoliosis, with data from patients with juvenile idiopathic arthritis, and with data from patients with migraine or cancer. Abbreviations, see Figure 1.

^a (Padua et al. 2001) Italy, $n = 70$ (54 F, 16 M), mean age 37 (26–49) years;

^b (Danielsson et al. 2001) Sweden, $n = 146$ (145 F, 11 M), mean age 40 (34–46) years;

^c (Foster et al. 2003) UK, $n = 82$ (68 F, 14 M), median age 30 (17–68) years;

^d (Aaronson et al. 1998) the Netherlands, $n = 423$ (84% F, 16% M), mean age 40 (16–88) years;

^e (Aaronson et al. 1998) the Netherlands, $n = 485$ (58% F, 42% M), mean age 57 (22–86) years.

of findings pertaining to the sample to the whole population. We consider that a 95% response is a good result.

Two limitations could not be avoided, however: the first one was intrinsic to the population analyzed, namely the inclusion criterion that patients had already been selected for operation. Thus, our findings are not readily applicable to the general population of common low back pain sufferers. The second limitation is the absence of a co-morbidity measure, which could help interpretation of the results of a generic HRQoL instrument such as SF-36. We used several techniques, such as the use of norm-based scoring, to overcome this potential limitation. By controlling for age and sex, some of the co-morbidity should be already taken into account.

Graphically (Figure 1), the differences from normative values as reported in the Swedish SF-36 manual (Sullivan et al. 1994) are striking, especially in the first 3 physical domains. They are lower than previously reported, even outside the normative studies both for LBP (McKinnon et al. 1997) and for other musculoskeletal conditions

(Stucki et al. 1995, Gartsman et al. 1998). The only comparable profile is the one published by Fanuele et al. (2000) from a large sample of patients from the US National Spine Network (NSN), a nonprofit organization whose members (academic institutions, hospitals, private physician practices, and individual physicians) treat patients with back and neck problems. The mean NSN values are more or less at the upper limit of our range of values—as can be expected since they do not include only surgical candidates, but a more general back pain population.

The low values of correlation coefficients between SF-36 domains and other preoperative clinical variables may be surprising. Especially regarding pain assessment, we would have expected stronger correlation between the Bodily Pain domain and assessment of back and leg pain on the VAS or the consumption of analgesics, such as the one between Physical Function and walking ability. Walsh et al. (2003) recently presented data that supported a hypothesis we had previously formulated (Zanoli et al. 2000)—that for studies of patients with low back problems, the general SF-36 might be a sufficient measure. We believe that our present findings reinforce the idea that a few additional, carefully chosen, variables (such as separate VAS assessment for Back and Leg Pain intensity) may add useful clinical information (Zanoli et al. 2001) which might not be included in the SF-36 scores, without the need for separate condition-specific instruments.

More recent clinical and commercial applications of the SF-family questionnaires introduced the use of norm-based scoring (Ware et al. 2000) to compare individual data with corresponding age- and sex-matched control group profiles. Our database provided only the traditional scores, so we performed a secondary calculation to obtain individual differences from expected normal values for each patient. The results are then presented as mean deviation from the norm for each diagnostic group; this allows a more immediate comparison (Figure 2) because the results already include adjustment for sex and age distribution in the sample. Unfortunately, not many reports include this measurement yet and it can only be calculated from the individual data, so few comparisons can be made. One exception comes from the musculoskeletal field

(Strömbeck et al. 2000). The authors presented this type of information in a group of patients with rheumatologic diagnoses (fibromyalgia, Sjögren's syndrome and rheumatoid arthritis). Sjögren's syndrome and rheumatoid arthritis patients were also significantly lower than the norm, but only fibromyalgia showed a very big gap in the first 3 physical domains, as in our sample. Interestingly, these rheumatological conditions also seemed to affect General Health scores and the mental domains in a rather homogeneous way, whereas in our sample the effect on physical domains was more instantly recognizable (Figure 4).

When this type of scoring is not available, comparison can be made graphically using the traditional SF-36 profile, which relies on mean values for each domain. In the examples presented in Figure 5, demographic data on the sample being compared are reported in the legend, to allow a more informed comparison. In all cases, values from our sample were strikingly lower than those published for other disabling conditions, although the absence of an age- and sex-adjustment makes the comparison less useful.

All the mean differences reported in Figure 2 can be considered to be statistically significant, because their confidence intervals never overlap the zero. Still, especially in the case of General Health, one can argue that some of these differences are not clinically relevant. It is always a major task to translate methodological research into useful clinical information (Fortin et al. 1995). We are not aware of any specific study dealing with Minimal Clinically Important Differences (MCIDs) of deviations from SF-36 normative values in LBP. One can speculate, taking inspiration from other methods of outcome assessment (Roland and Morris, Oswestry, VAS), that 10–20% of the highest possible value should represent a clinical difference: in our case, it would mean that a score which has a mean difference of more than 20 points from the norm probably represents a population which is clinically different from the normal population regarding that particular SF-36 domain. In other words, we can say that surgical degenerative lumbar spine disorders affected HRQoL in our sample in a clinically significant way in all SF-36 domains except General Health (GH) and Mental Health (MH) (Figure 4, square markers).

Patients selected for surgery appear to be rather different from healthy people or nonspecific LBP patients, but their profiles appear rather homogeneous from one diagnostic group to the other. This happens despite the expected difference in median age between groups (Figure 3) (Zanoli et al. 2001). Only a few SF-36 domains, i.e. Physical Function, Bodily Pain and Social Function, showed some degree of differentiation between groups when plotted together with 95% CI (Figure 3).

The most clear-cut distinction occurs in the Bodily Pain domain, where lumbar disk herniation (HD) patients seem to experience significantly more pain than the others. We have previously described a similar pattern for separate VAS assessment of Back and Leg Pain Intensity (Zanoli et al. 2001). Some discrimination can also be made as regards Physical Function (with central stenosis and degenerative disk disease (DDD) patients being worse than the rest) and Social Function (with DDD patients being worse than the rest).

In contrast to the report by Vogt et al. (2002) on a multicenter, cross-sectional analysis of data from the National Spine Network, which is a much larger database but includes a broad spectrum of spinal conditions and different degrees of severity, smoking habits had no effect on perceived HRQoL. This might also be an effect of sample size, although neither of these studies was experimentally powered to detect any a priori clinically meaningful difference. Similarly, diagnostic workout did not seem to be influenced by the patients' HRQoL as assessed by SF-36: scores were similar for patients who underwent a minimal diagnostic pathway and those who needed a higher number of preoperative tests. Statistically significant differences could be ascertained in physical domains and in the RE domain when taking into account physical requirement at work, or splitting the sample in terms of workers and non-workers (on sick leave, unemployed, or retired). It is worth emphasizing that long-term worker's compensation status does not affect HRQoL as much as an acute sick leave. This latter finding is strikingly consistent with another report from the National Spine Network on neck pain patients (Hee et al. 2002).

The main feature of the graphs is the presence of a superimposable pattern in SF-36 profiles of our sample, which is constant in all the possible

Table 3. General objective criteria for the 5 diagnostic groups. These criteria were set a posteriori to help the interpretation of the data, and should not be regarded as strict and cogent definitions. When not otherwise specified, symptoms should normally last at least 6–8 weeks despite appropriate nonoperative treatment

Diagnosis	Criteria for surgical indication
Herniated disk	Radiating pain in a single dermatome and/or neuromuscular deficits at the same level. MRI or CT confirmation.
Central stenosis	Neurogenic claudication relieved in flexion. MRI or CT confirmation.
Degenerative/segmental pathology	Chronic mechanical LBP resistant to conservative treatment, with or without radiating pain. 1 of the 2 should last at least 6 months. MRI or CT confirmation.
Spondylolisthesis	Radiating pain in a single dermatome and/or neuromuscular deficits at the same level. Conventional radiography and MRI or CT confirmation.
Lateral stenosis	Radiating pain in a single dermatome and/or neuromuscular deficits at the same level. MRI or CT confirmation.

subsets of patients. One possible explanation for this finding is that the strong selection made before setting an indication for surgery (Table 3) probably succeeded in choosing a group of patients whose perceived HRQoL was somewhat similar, and so much influenced by the disease being treated that other variables became irrelevant.

Figure 3 highlights several significant diagnosis-specific differences in some domains, within the wide range of values observed: for example, degenerative disk disease patients appear to have more pain and spondylolisthesis patients appear to be less socially affected than others. It should be emphasized that this is not a specially selected sample, but it represents "a real life scenario" and is representative of all the patients who are treated in a specialized spinal unit in a large university hospital. Cases operated here are possibly more severe than those operated in a non-teaching hospital (Rosenthal et al. 1997). Duration of symptoms before the operation certainly represents the effect of long waiting lists rather than the application of

time-related selection criteria or non-surgical treatment methods.

Very often, especially at multidisciplinary meetings, physicians with different specialty training have difficulty in agreeing on the best treatment for a particular group of patients, because they do not agree on the diagnosis which might describe a specific clinical presentation. Surgeons may believe that there is no need to use the term fibromyalgia for patients who simply have (psychological) pain everywhere. Physiotherapists and rehabilitation specialists could argue that most adults have degenerative disk disease without requiring surgical intervention, and anyone who has not had a special subspecialty training might have difficulty in accepting the centralization phenomenon or the existence of tooth-related low back pain.

There is probably some truth in all of these statements. Still, in the group of surgically selected people we have studied, we have been able to describe a rather homogeneous pattern of perceived HRQoL—which supports the hypothesis that they actually differ from other low back pain patients. The higher impairment in physical function caused by central stenosis and degenerative disc disease, the more pronounced symptoms in herniated nucleus pulposus and the social limitations of the somewhat more problematic degenerative disk disease patients are also clinically consistent. We are not proposing SF-36 as a diagnostic tool; nor are we pretending that our limited findings can be generalized. In this respect, one could also note that age of the patients at presentation also shows a similar discriminative ability. We rather prefer to interpret this finding as an implicit confirmation of the selection process of surgical patients in our clinic; for example, one would expect lumbar disk herniation patients to be younger than central stenosis patients, and this is confirmed by our observation. Actually, confirming diagnoses was not the aim of the study, and the protocol was not specifically designed to discriminate between surgical and non-surgical patients, so we cannot draw any conclusions. However, from a practical point of view, this clinical consistency is probably the most rewarding information yielded by the data analysis of our sample.

Author contributions

GZ designed study, elaborated and analysed data, wrote draft manuscript, BJ initiated data collection, collected data, commented draft manuscript, BS initiated data collection, designed study, collected data, corrected draft manuscript.

This study was supported by the Swedish National Board of Health and Welfare, the Medical Faculty of Lund University, the Foundation for Support of the Disabled in Skåne, and the Swedish Science Research Council (grant 09509).

No competing interests declared.

Aaronson N K, Muller M, Cohen P D, Essink-Bot M L, Fekkes M, Sanderman R, Sprangers M A, te Velde A, Verrips E. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998; 51 (11): 1055-68.

Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders. Introduction. *Spine* 2000a; 25 (24): 3097-9.

Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 2000b; 25 (24): 3100-3.

Conover W J. Practical Nonparametric Statistics. John Wiley and Sons, New York, 1980; cit. in Bland M. An introduction to medical statistics. Oxford University Press, 1980.

Coulter I D, Hurwitz E L, Adams A H, Genovese B J, Hays R, Shekelle P G. Patients using chiropractors in North America: who are they, and why are they in chiropractic care? *Spine* 2002; 27 (3): 291-6.

Dahl B, Gehrchen P M, Kiaer T, Blyme P, Tondenvold E, Bendix T. Nonorganic pain drawings are associated with low psychological scores on the preoperative SF-36 questionnaire in patients with chronic low back pain. *Eur Spine J* 2001; 10 (3): 211-4.

Danielsson A J, Wiklund I, Pehrsson K, Nachemson A L. Health-related quality of life in patients with adolescent idiopathic scoliosis: a matched follow-up at least 20 years after treatment with brace or surgery. *Eur Spine J* 2001; 10 (4): 278-88.

Fanuele J C, Birkmeyer N J, Abdu W A, Tosteson T D, Weinstein J N. The impact of spinal problems on the health status of patients: have we underestimated the effect? *Spine* 2000; 25 (12): 1509-14.

Fortin P R, Stucki G, Katz J N. Measuring relevant change: an emerging challenge in rheumatologic clinical trials. *Arthritis Rheum* 1995; 38 (8): 1027-30.

Foster H E, Marshall N, Myers A, Dunkley P, Griffiths I D. Outcome in adults with juvenile idiopathic arthritis: a quality of life study. *Arthritis Rheum* 2003; 48 (3): 767-75.

Gartsman G M, Khan M, Hammerman S M. Arthroscopic repair of full-thickness tears of the rotator cuff. *J Bone Joint Surg (Am)* 1998; 80 (6): 832-40.

- Hee H T III, Whitecloud T S 3rd, Myers L, Roesch W, Ricciardi J E. Do worker's compensation patients with neck pain have lower SF-36 scores? *Eur Spine J* 2002; 11 (4): 375-81.
- McKinnon M E, Vickers M R, Ruddock V M, Townsend J, Meade T W. Community studies of the health service implications of low back pain. *Spine* 1997; 22 (18): 2161-6.
- Padua R, Padua S, Aulisa L, Ceccarelli E, Padua L, Romanini E, Zanolì G, Campi A. Patient outcomes after Harrington instrumentation for idiopathic scoliosis: a 15- to 28-year evaluation. *Spine* 2001; 26 (11): 1268-73.
- Rosenthal G E, Harper D L, Quinn L M, Cooper G S. Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals. Results of a regional study. *JAMA* 1997; 278 (6): 485-90.
- Stewart A L, Ware J E. Measuring functioning and well-being: The medical outcomes study approach. Durham N C: Duke University Press, 1992
- Strömbeck B, Ekdahl C, Manthorpe R, Wikström I, Jacobsson L. Health-related quality of life in primary Sjögren's syndrome, rheumatoid arthritis and fibromyalgia compared to normal population data using SF-36. *Scand J Rheumatol* 2000; 29 (1): 20-8.
- Strömqvist B, Jönsson B. Computerized follow-up after surgery for degenerative lumbar spine diseases. *Acta Orthop Scand (Suppl 251)* 1993:138-42.
- Strömqvist B, Jonsson B, Fritzell P, Hägg O, Larsson B E, Lind B. The Swedish National Register for lumbar spine surgery: Swedish Society for Spinal Surgery. *Acta Orthop Scand* 2001; 72 (2): 99-106.
- Stucki G, Liang M H, Phillips C, Katz J N. The Short Form-36 is preferable to the SIP as a generic health status measure in patients undergoing elective total hip arthroplasty. *Arthritis Care Res* 1995; 8 (3): 174-81.
- Sullivan M, Karlsson J, Ware J E, Hälsoenkät SF-36. Svensk manual och tolkningsguide (SF-36 Health Survey. Swedish Manual and Interpretation Guide), Sahlgrenska University Hospital, Göteborg, 1994.
- Vogt M T, Hanscom B, Lauerma W C, Kang J D. Influence of smoking on the health status of spinal patients: the National Spine Network database. *Spine* 2002; 27 (3): 313-9.
- Walsh T L, Hanscom B, Lurie J D, Weinstein J N. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine* 2003; 28 (6): 607-15.
- Ware J E Jr, Sherbourne C D. The MOS 36-item Short-Form health survey (SF-36). 1. Conceptual framework and item selection. *Med Care* 1992; 30: 473-83.
- Ware J E Jr, Snow K K, Kosinski M, Gandek B. SF-36 Health Survey: Manual and Interpretation guide, New England Medical Center, The health Institute, Boston, MA, 1993.
- Ware J E Jr, Björner J B, Kosinski M. Practical implications of item response theory and computerized adaptive testing: a brief summary of ongoing studies of widely used headache impact scales. *Med Care (Suppl 9)* 2000; 38: II73-82.
- Zanolì G, Strömqvist B, Padua R, Romanini E. Lessons learned searching for a HRQoL instrument to assess the results of treatment in persons with lumbar disorders. *Spine* 2000; 25 (24): 3178-85.
- Zanolì G, Strömqvist B, Jönsson B. Visual analog scales for interpretation of back and leg pain intensity in patients operated for degenerative lumbar spine disorders. *Spine* 2001; 26 (21): 2375-80.