

Reliability of the prospective data collection protocol of the Swedish Spine Register

Test-retest analysis of 119 patients

Gustavo Zanolini^{1,2}, Lars T Nilsson¹ and Björn Strömqvist¹

Departments of Orthopedics, ¹Lund University Hospital, SE-221 85 Lund, Sweden, ²University of Ferrara, IT-44100 Ferrara, Italy
Correspondence GZ: gzanoli@globeweb.org
Submitted 05-01.30 Accepted 06-02-04

Background The Swedish Lumbar Spine Register has been collecting patient-based data since 2000, and more than 80% of all spinal units in Sweden are now including their patients. In a few years, it will produce useful clinical information just as arthroplasty registers have, but to permit proper interpretation of data in the future, the reliability of the protocol must be tested.

Methods Between January 2000 and March 2003, a sample of 122 patients was asked to fill in the questionnaire twice: 63 preoperatively and 59 postoperatively. Test-retest reliability was calculated with intra-class correlation coefficient (ICC) or weighted kappa when appropriate.

Results Test-retest interval varied (range 0–235 days); in the “worst case scenario”, the lowest ICC for SF-36 was 0.62 for the postoperative RE. Other values were above 0.70; for non-SF variables, ICC was in the range 0.79–0.89. Kappa values for the ordinal outcomes were high (0.74–0.91).

Interpretation When separate reliability analysis was performed according to the time interval, a 0–2 days interval produced a significant memory effect; after 3 weeks, the reliability seemed to drop in the preoperative group, whereas results were reproducible up to 9 weeks postoperatively. The protocol studied can reliably detect postoperative improvements between large groups of patients such as in a register.

reorganization in 1998 that data collection began in a consistent way on a national basis. Changes introduced at that time included administrative and computer support for participating departments, and also the adoption of a patient-based protocol of data collection (Strömqvist et al. 2001) which simplified the routines for registering surgeons and played a major role in achieving satisfactory national coverage, which is crucial for the validity of any register.

The choice of patient-oriented outcomes was made mainly from scientific considerations, since meantime psychometrics (Deyo et al. 1994, Hägg 2002) and, in the clinical field, the outcomes movement (Swiontkowski et al. 1999) had already gone a long way to proving the validity of self-administered questionnaires and standardization methods, producing an even too large number of different questionnaires (Zanolini et al. 2000).

Among other demographic variables and questions on local pain, radiating pain, analgesic intake, walking ability and patient satisfaction, the SF-36 health survey (Ware and Sherbourne 1992) was chosen to evaluate health-related quality of life (HRQoL). Even though this instrument has been extensively validated and can be considered to be one of the standards of outcomes measurement for spinal disorders (Deyo 1998, Bombardier 2000, Ware 2000, Walsh et al. 2003), its inclusion in a more comprehensive data collection protocol (Hunsaker et al. 2002) and use in a specific clinical setting (Hanscom et al. 2002) usually requires some

The Swedish Lumbar Spine Register was officially started in 1993, but it was only after a major

form of standardization and re-evaluation (Streiner and Norman 1995, Zanoli 2001) of the new protocol. We have already addressed the issues of normative standardization (Zanoli et al. 2002) and responsiveness (Zanoli et al. 2005) in this group of patients. In the present study, we evaluated test-retest reliability of the prospective data collection protocol adopted by the Swedish Lumbar Spine Register in a sample of patients at our institution.

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 Register. Details of the data collection protocol and patient selection have been reported elsewhere (Strömqvist et al. 2001). More information and downloadable forms are available at <http://www.4s.nu>. In addition to SF-36 questionnaire responses, separate local and radiating pain (intensity on a VAS, duration in months), change in pain, analgesic intake, working status, walking ability and patient satisfaction were recorded, together with other variables not included in the present analysis (age, sex, duration of sickleave, number of previous operations, smoking habits, number and type of preoperative diagnostic procedures performed). The patients completed the preoperative questionnaires independently on the day of admission for surgery and after 1 year postoperatively during a follow-up visit or, if no visit was planned, by mail using a prepaid envelope. Demographic variables and duration of symptoms were recorded only preoperatively. Pre- and postoperative VAS scores were obtained in standard fashion, i.e. by measuring the distance in mm from the origin of a horizontal line (total length 100 mm) and the point indicated by the patient as representing his/her level of pain during the previous week. Zero represented “no pain at all” and 100 represented “the worst pain imaginable”. Analgesic intake was recorded on a 3-point Likert Scale (regular, intermittent, none). Walking distance was recorded on a 4-point Likert Scale (less than 100 m, 100–500 m, 500 m–1 km, more than 1 km).

Postoperatively, changes in leg and back pain were recorded on a 5-point Likert Scale (Pain-free, Much Better, Somewhat Better, Unchanged, Worse) as compared to preoperative status. The patient's working status was recorded on a 6-point Likert Scale (Back to previous activity, Back to lighter activity, Sick leave because of back, Sick leave because of other problems, Unemployed, Retired). Patient satisfaction was recorded on a 3-point Likert Scale (Satisfied, Uncertain, Dissatisfied).

To the best of our knowledge, this procedure is standard for all units in Sweden.

Between November 2000 and March 2003, a sample of 122 patients was asked to fill in the questionnaire twice. 63 (median age 55 (26–84) years, 34 men) were given it at the time of the last preoperative visit (in order for us to be able to compare the results with the standard preoperative questionnaire) and they were asked to fill in the questionnaire at home 1 week before coming to the hospital and to bring it when being admitted to hospital. 59 (median age 53 (22–81) years, 30 women) received a second copy at the time of the 1-year follow-up visit (in order for us to be able to compare the results with those of the standard 1-year follow-up) and were asked to fill in the questionnaire at home after 1 week and then to return it to the department by mail. The patients could only belong to one of the 2 groups, depending on what stage of their registration they were cross-sectionally during the study period. In the same period, 302 other patients were registered; a Mann-Whitney test was performed to ensure that there were no significant differences between the sample and the whole population regarding demographics and preoperative variables. Diagnoses in the sample were distributed as follows (herniated disc = 35, central stenosis = 28, degenerative/segmental pathology = 17, spondylolisthesis = 12, lateral stenosis = 28, other/combination = 2). This distribution is comparable to that observed in previously published material (Zanoli et al. 2006) and was representative of the present selection of patients at our institution. For a more detailed description of the diagnostic criteria used, see also (Zanoli et al. 2003). In all cases, the date of completion of the questionnaire was entered by the patients on the same forms.

Table 1. Interval between test and retest in the preoperative and postoperative sample

Sample:	Preoperative n=60	Postoperative n=59	Total n=119
Days between test and retest			
Minimum	0	3	0
Maximum	235	118	235
Median	12	32	24

Test-retest reliability for continuous variables (VAS, SF-36 scores) was calculated using a two-way random model intraclass correlation coefficient (ICC) for each score as the ratio of variance between subjects and the total variance (Streiner and Norman 1995, Nichols 1998). ICCs range from 0 (least reliable) to 1 (most reliable). Weighted kappa calculation (Cohen 1968) was used to assess agreement between the first and the second survey for ordinal variables such as change in pain, analgesic intake, working status, walking ability and patient satisfaction. Kappa also ranges from 0 (worst agreement) to 1 (best agreement). For two SF-36 scores (RP and RE) which present continuous values but actually have only a limited number of possible levels of answers (7 for RP and 5 for RE), both ICCs and kappas were calculated.

Data collection and analysis

Data were collected in a personalized Filemaker Pro database (Knutson 1993–2002) using Microsoft Excel 2002. Statistical analysis was performed by using SPSS for Windows version 11.5. MedCalc 7.6.0.0 (www.medcalc.be) was used for calculation of weighted kappa.

Results

3 of 63 patients in the preoperative sample were not included because they did not return the questionnaire. After examining the self-entered date of completion in the returned forms, we realized that not all patients had complied with the 1-week recommendation, but had answered the questionnaire after different lengths of time (Table 1). Thus, we decided to run separate reliability analyses for different subgroups, which were established post hoc, as this scattered response was not expected. The rationale for group categorization was to exclude memory effect (at least 3-day interval) and to explore the effect of time as if it had been an inclusion criterion of the study, thus excluding all the patients who did not fit in the specified interval (Table 2). Cutpoints are absolutely arbitrary and should be considered as such in the interpretation.

No significant difference (Mann-Whitney test, $p < 0.05$) was found regarding demographics and preoperative variables between the patients who were included and other patients registered in the same period, and no significant difference was found between subgroups.

Reliability results for the whole sample and the different time-related subgroups are presented in Tables 3–6. Apart from providing the necessary data to fulfill the aim of our study, we believe that providing the reliability coefficients for the different subgroups may be a useful source of reference for future studies on reliability or outcome studies in similar clinical settings.

In the “worst case scenario”, i.e. analyzing the total material irrespective of time elapsed between

Table 2. Interval between test and retest in the 5 preoperative and 5 postoperative subgroups

Subgroups: ^a	Preoperative sample					Postoperative sample				
	3–63d n=44	3–19d n=30	3–30d n=42	20–63d n=14	0–2d n=10	3–63d n=54	3–19d n=9	3–30d n=24	20–63d n=45	0–2d n=0
Days between test and retest										
Minimum	4	4	4	20	0	3	3	3	20	–
Maximum	46	19	30	46	1	63	19	30	63	–
Median	12	10	12	24	0	31	10	25	33	–

^a Subgroups after time interval in days (3–63d = time interval between 3 and 63 days; 3–19d = time interval between 3 and 19 days; 3–30d = time interval between 3 and 30 days; 20–63d = time interval between 20 and 63 days; 0–2d = time interval between 0 and 2 days)

Table 3. Intraclass correlation coefficients for the outcome variables in the preoperative sample

Variables ^a	Whole sample		Pre 3–63d		Pre 3–19d		Pre 3–30d		Pre 20–63d		Pre 0–2d	
	n	ICC ^b	n	ICC ^a	n	ICC ^a	n	ICC ^a	n	ICC ^a	n	ICC ^a
Back pain (months)	60	0.85	44	0.84	30	0.87	42	0.89	14	0.80	10	1.00
95% CI		0.77–0.91		0.73–0.91		0.74–0.93		0.80–0.94		0.50–0.93		–
Leg pain (months)	60	0.84	44	0.81	30	0.92	42	0.81	14	0.08	10	0.99
95% CI		0.75–0.90		0.67–0.89		0.85–0.96		0.67–0.89		–0.46–0.57		0.98–1.00
Back pain (VAS)	60	0.78	44	0.74	30	0.85	42	0.77	14	0.43	10	1.00
95% CI		0.66–0.87		0.56–0.85		0.70–0.92		0.61–0.87		–0.14–0.78		–
Leg pain (VAS)	60	0.88	44	0.89	30	0.93	42	0.88	14	0.80	10	1.00
95% CI		0.81–0.93		0.80–0.94		0.85–0.96		0.79–0.94		0.50–0.93		–
PF	60	0.82	44	0.81	30	0.85	42	0.82	14	0.75	10	1.00
95% CI		0.72–0.89		0.67–0.89		0.70–0.93		0.70–0.90		0.40–0.91		–
RP	60	0.85	44	0.79	30	0.79	42	0.79	14	0.77	10	1.00
95% CI		0.76–0.91		0.65–0.88		0.60–0.89		0.64–0.88		0.44–0.92		–
BP	60	0.69	44	0.64	30	0.67	42	0.64	14	0.61	10	0.87
95% CI		0.51–0.80		0.38–0.80		0.34–0.84		0.38–0.80		0.17–0.85		0.57–0.96
GH	60	0.86	44	0.87	30	0.85	42	0.86	14	0.91	10	0.91
95% CI		0.77–0.92		0.72–0.93		0.69–0.93		0.72–0.93		0.60–0.97		0.71–0.98
VT	60	0.67	44	0.62	30	0.64	42	0.62	14	0.60	10	0.87
95% CI		0.46–0.80		0.34–0.79		0.27–0.83		0.34–0.79		0.15–0.85		0.57–0.96
SF	60	0.77	44	0.76	30	0.86	42	0.75	14	0.35	10	0.72
95% CI		0.65–0.86		0.60–0.86		0.74–0.93		0.58–0.86		–0.21–0.73		0.25–0.92
RE	60	0.67	44	0.62	30	0.74	42	0.59	14	0.36	10	1.00
95% CI		0.51–0.79		0.40–0.77		0.53–0.87		0.36–0.76		–0.21–0.74		–
MH	57	0.89	42	0.88	29	0.89	40	0.88	13	0.84	9	1.00
95% CI		0.83–0.94		0.78–0.93		0.79–0.95		0.78–0.93		0.54–0.95		–

^a Variables: Back pain (months) = Preoperative duration of back pain in months.
 Leg pain (months) = Preoperative duration of leg pain in months.
 Back pain (VAS) = Preoperative VAS for back pain intensity.
 Leg pain (VAS) = Preoperative VAS for leg pain intensity.
 Preoperative SF–36 domains: PF = Physical Function, RP = Role Physical, BP = Bodily Pain, GH = General Health,
 VT = Vitality, SF = Social Function, RE = Role Emotional, and MH = Mental Health,
 95% CI = 95% confidence interval of ICC.
^b ICC = two-way random model Intraclass Correlation Coefficient.

Table 4. Intraclass correlation coefficients for the outcome variables in the postoperative sample

Variable	Whole sample			Post 3–63d			Postoperative subgroups (see Table 2)			Post 20–63d					
	n	ICC	95%CI	n	ICC	95%CI	n	ICC	95%CI	n	ICC	95%CI			
Back pain ^a	55	0.89	0.82–0.93	52	0.90	0.83–0.94	9	0.98	0.91–0.99	24	0.89	0.76–0.95	43	0.88	0.79–0.93
Leg pain ^a	55	0.85	0.75–0.91	52	0.87	0.78–0.92	9	0.90	0.63–0.98	24	0.91	0.80–0.96	43	0.86	0.75–0.93
PF	59	0.86	0.78–0.92	54	0.91	0.84–0.94	9	0.93	0.75–0.98	24	0.93	0.85–0.97	45	0.90	0.83–0.94
RP	59	0.71	0.56–0.82	54	0.71	0.55–0.82	9	0.66	0.11–0.91	24	0.76	0.52–0.89	45	0.73	0.56–0.84
BP	59	0.86	0.78–0.92	54	0.89	0.82–0.94	9	0.84	0.45–0.96	24	0.87	0.72–0.94	45	0.91	0.83–0.95
GH	59	0.80	0.68–0.88	54	0.84	0.70–0.91	9	0.91	0.67–0.98	24	0.91	0.79–0.96	45	0.82	0.67–0.91
VT	59	0.87	0.79–0.92	54	0.88	0.81–0.93	9	0.93	0.71–0.98	24	0.88	0.74–0.95	45	0.88	0.79–0.93
SF	59	0.65	0.48–0.78	54	0.71	0.54–0.82	9	0.69	0.08–0.92	24	0.67	0.36–0.84	45	0.72	0.54–0.84
RE	59	0.62	0.43–0.76	54	0.62	0.42–0.76	9	0.31	0.45–0.79	24	0.65	0.34–0.83	45	0.69	0.50–0.82
MH	59	0.83	0.73–0.89	54	0.85	0.76–0.91	9	0.84	0.48–0.96	24	0.86	0.67–0.94	45	0.85	0.75–0.92

For abbreviations, see Table 3
^a Postoperative VAS for pain intensity

Table 5. Agreement (weighted kappa) for the outcome variables in the preoperative sample

Variable ^a	Whole sample			Pre 3–63d			Preoperative subgroups (see Table 2)										
	n	κ	SE	n	κ	SE	Pre 3–19d			Pre 3–30d			Pre 20–63d			Pre 0–2d	
Work	57	0.91	0.04	42	0.93	0.04	28	0.93	0.04	40	0.93	0.04	14	0.95	0.05	9	1.00
Analg.	59	0.85	0.06	43	0.80	0.08	30	0.85	0.09	41	0.78	0.08	13	0.69	0.16	10	1.00
Walk	58	0.81	0.05	42	0.78	0.07	29	0.70	0.10	41	0.77	0.07	13	0.93	0.07	10	1.00
RP	60	0.76	0.08	44	0.67	0.11	30	0.65	0.12	42	0.67	0.11	14	0.76	0.23	10	1.00
RE	60	0.64	0.10	44	0.57	0.12	30	0.68	0.12	42	0.55	0.12	14	0.34	0.23	10	1.00

^a Variables: n = number of patients, κ = weighted kappa, SE = standard error.

Work = Preoperative working situation.

Analg. = Preoperative use of pain-killers.

Walk = Preoperative walking distance.

RP = Preoperative SF-36 – Role Physical.

RE = Preoperative SF-36 – Role Emotional.

Table 6. Agreement (weighted kappa) for the outcome variables in the postoperative sample

Variable ^a	Whole sample			Post 3–63d			Postoperative subgroups (see Table 2)								
	n	κ	SE	n	κ	SE	Post 3–19d			Post 3–30d			Post 20–63d		
Back pain change ^b	54	0.74	0.06	51	0.74	0.06	8	0.66	0.19	23	0.72	0.11	43	0.75	0.07
Leg pain change ^b	53	0.85	0.04	50	0.86	0.04	8	1.00	–	22	0.94	0.05	42	0.83	0.05
Work	55	0.91	0.04	52	0.93	0.04	9	1.00	–	24	0.98	0.02	43	0.91	0.05
Analg.	53	0.87	0.05	50	0.89	0.05	9	0.87	0.13	24	0.85	0.09	41	0.89	0.05
Walk	55	0.77	0.07	52	0.77	0.07	9	0.65	0.22	24	0.78	0.11	43	0.80	0.06
Satisfaction	51	0.83	0.09	48	0.91	0.06	8	1.00	–	21	0.92	0.09	40	0.89	0.08
RP	59	0.64	0.08	54	0.64	0.08	9	0.56	0.21	24	0.70	0.12	45	0.65	0.09
RE	59	0.53	0.09	54	0.54	0.10	9	0.27	0.27	24	0.57	0.14	45	0.59	0.10

^a Variables: n = number of patients, κ = weighted kappa, SE = standard error.

^b 12 months self-rated pain improvement

Work = 12 months working situation.

Analg. = 12 months use of pain-killers.

Walk = 12 months walking distance.

Satisfaction = 12 months patient's satisfaction.

RP = 12 months SF-36 – Role Physical.

RE = 12 months SF-36 – Role Emotional.

completion of questionnaires, the lowest ICC for SF-36 scores was 0.62 for the RE scale in the postoperative group, which also showed only “moderate” agreement using kappa values. Most other kappa values were above 0.70, and for non-SF variables they were in the range 0.79–0.89 (Tables 3–6). Kappa values for the other ordinal outcome questions were quite high (0.74–0.91) (Tables 5–6).

Discussion

The Swedish Lumbar Spine Register protocol was designed to achieve discrimination between groups, possibly large groups of patients. Our practical conclusion from the analysis of the present data and comparison with similar data in the literature (Roos et al. 1998, Dunbar et al. 2001, Hagberg et al. 2004, Ostelo et al. 2004) is that all the outcome variables under study appear to be sufficiently reliable (Bland and Altman 1986, 2002, Altman 1991)—although some might be more so than

others—to detect the rather pronounced improvement which is usually seen postoperatively in this type of patient (Zanoli et al. 2005).

To be more speculative, looking at the different tables, we could say that to allow good generalizability of the results, postoperative measurement can be performed within a range of ± 2 months and still be reliable, whereas preoperative values within 2–3 weeks seem to yield more reproducible information. It is interesting to note that confidence intervals became too wide to draw any conclusion when the sample size was less than 20, with the exception of subgroup PRE0-2d, where the extremely high correlation was probably due to a memory effect because of the short between the two surveys, as confirmed by the absolute agreement seen on the kappa values. Conversely, when the time span was more than 20 days, reliability dropped for many variables in the Pre 20–63d group. This did not seem to affect the postoperative measurement as much, as it can be seen in the Post 20–63d group. Choosing the one month time distance in our sample would probably represent a good compromise between stability over time (which increases the values of coefficients) and the number of subjects studied (which narrows the confidence intervals).

Apart from these considerations, we cannot really indicate an ideal time span from our data, so we have chosen to present the data and to leave it up to the reader to choose, depending on the specific requirements. In general, postoperative results seemed to be more stable and less affected by a longer time span between the two measurements. This may be due to the fact that the improvement that can be achieved—which we have shown elsewhere (Zanoli et al. 2005)—is not prone to the day-to-day modifications that the wavering preoperative symptoms of LBP are (Maul et al. 2003). It is notable that separate VAS measurements of back and leg pain seem to yield more reliable results than the bodily pain subscale of the SF-36. This is probably due to the fact that with this particular type of pathology, back pain and leg pain are distinct entities and not differentiating between the two (as in the SF-36 pain questions) may produce less reliable information.

No attempt was made to characterize reliability within subsets of patients with the same diagnosis

or who underwent the same surgical intervention. Even if it is possible that this would have led to better figures, a much larger sample would have been required and this was beyond the scope of the present study. The participation rate in the Swedish Lumbar Spine Register has been increasing since 2000, and 80% of all spinal units in Sweden are now submitting patient details (Strömqvist et al. 2001). Hopefully, in a few years it will be able to produce useful clinical information, just as arthroplasty registers have (Herberts and Malchau 2000). The register became internet-based in January 2003, and two more questionnaires have been added to the protocol: the Oswestry disability score and the EuroQol. As they were not part of the original protocol when this project started, these two questionnaires have not been considered in the present study, and should be explored in future studies. However, despite sharing the enormous practical difficulties of running a large multicenter nationwide data collection project, Spine and Arthroplasty Registers have little in common. Patient description and outcome evaluation in spinal surgery cannot rely on simple categories (trauma, degenerative or inflammatory disease) and dichotomous outcomes (revision). This has many obvious drawbacks (more time consuming data collection and reduced compliance). However, this is also the reason why spine surgeons have been among the first to lead the way to patient-centered evaluation.

All the scientific knowledge accumulated can become useful in today's hard competition for resources, when the effectiveness and cost-effectiveness of many spinal surgeries is questioned. Carefully designed randomized controlled trials (RCTs) are needed to reach valid conclusions on effectiveness or cost-effectiveness. Registers are not a substitute for RCTs, as they do not control for a number of factors (placebo effect, natural course, concomitant therapies etc.). However, registers can be very useful for gathering data on safety, for generating data supporting the need—and opportunity—for conducting randomized controlled trials, and for providing data that can help in the design of such RCTs and in assessing their feasibility. In addition, they can also provide very useful information for comparison of results in routine clinical practice (“audit” through “registers”) with those

obtained under more experimental conditions (RCTs on efficacy or effectiveness). Although the degree of recovery that is directly attributable to surgical treatment cannot be determined with observational data such as those obtained from register studies, we hope that the information provided in this article will increase the validity and interpretability of the results of the Swedish Lumbar Spine Register and similar studies.

Contributions of authors

GZ designed the study, elaborated and analyzed data, and wrote the draft manuscript. LTN collected data, commented on the study design, and commented on the draft manuscript. BS started data collection, designed the study, collected data, and corrected the draft manuscript.

We wish to thank Jonas Ranstam for statistical advice. Sources of funding were the Swedish National Board of Health and Welfare, the Medical Faculty of Lund University, the Swedish Science Research Council (grant 09509), and the Foundation for Support of the Physically Handicapped in Skåne.

- Altman D G. Practical statistics for medical research. London: Chapman and Hall, 1991.
- Bland J M, Altman D G. Statistical method for assessing agreement between two methods of clinical measurement. *The Lancet* 1986; 1: 307-10.
- Bland J M, Altman D G. Statistics Notes: Validating scales and indexes. *BMJ* 2002; 324 (7337): 606-7.
- Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 2000; 25 (24): 3100-3.
- Cohen J. Weighted kappa: nominal scale agreement with provision for scale and disagreement or partial credit. *Psychol Bull* 1968, 70: 213-20.
- Deyo R A, Andersson G, Bombardier C, Cherkin D C, Keller R B, Lee C K, Liang M H, Lipscomb B, Shekelle P, Spratt K F, et al. Outcome measures for studying patients with low back pain. *Spine (Suppl 18)* 1994; 19: 2032S-6S.
- Deyo R A, Battie M, Beurskens A J, Bombardier C, Croft P, Koes B, Malmivaara A, Roland M, Von Korff M, Waddell G. Outcome measures for low back pain research. A proposal for standardized use. *Spine* 1998; 23 (18): 2003-13.
- 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.
- Hagberg K, Branemark R, Hägg O. Questionnaire for Persons with a Transfemoral Amputation (Q-TFA): Initial validity and reliability of a new outcome measure. *J Rehabil Res Dev* 2004; 41 (5): 695-706.
- Hägg O. Measurement and prediction of outcome. Application in fusion surgery for chronic low-back pain. Göteborg University, Thesis, Göteborg, Sweden 2002
- Hanscom B, Lurie J D, Homa K, Weinstein J N. Computerized questionnaires and the quality of survey data. *Spine* 2002; 27 (16): 1797-801.
- Herberts P, Malchau H. Long-term registration has improved the quality of hip replacement: a review of the Swedish THR Register comparing 160,000 cases. *Acta Orthop Scand* 2000; 71 (2): 111-21.
- Hunsaker F G, Cioffi D A, Amadio P C, Wright J G, Caughlin B. The American academy of orthopaedic surgeons outcomes instruments: normative values from the general population. *J Bone Joint Surg (Am)* 2002; 84 (2): 208-15.
- Maul I, Laubli T, Klipstein A, Krueger H. Course of low back pain among nurses: a longitudinal study across eight years. *Occup Environ Med* 2003; 60 (7): 497-503.
- Nichols D P. Choosing an intraclass correlation coefficient. SPSS Keywords, Number 67, 1998. Available from: URL: <http://www.ats.ucla.edu/stat/spss/library/whichicc.htm>.
- Ostelo R W, de Vet H C, Knol D L, van den Brandt P A. 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *J Clin Epidemiol* 2004; 57 (3): 268-76.
- Roos E M, Roos H P, Lohmander L S, Ekdahl C, Beynon B D. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. *J Orthop Sports Phys Ther* 1998; 28 (2): 88-96.
- Streiner D L, Norman G R. Health measurement scales: a practical guide to their development and use. 2nd ed. Oxford: Oxford University Press, 1995.
- Strömquist B, Jönsson B. Computerized follow-up after surgery for degenerative lumbar spine diseases. *Acta Orthop Scand (Suppl 251)* 1993: 138-42.
- Strömquist B, Jönsson B, Fritzell P, Hägg O, Larsson B-E, Lind B. The Swedish national register for lumbar spine surgery. *Acta Orthop Scand* 2001; 72: 99-106.
- Swiontkowski M F, Buckwalter J A, Keller R B, Haralson R. The outcomes movement in orthopaedic surgery: where we are and where we should go. *J Bone Joint Surg (Am)* 1999; 81 (5): 732-40.
- 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. SF-36 health survey update. *Spine* 2000; 25 (24): 3130-9.
- 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.

- Zanoli 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.
- Zanoli G, Strömqvist B, Jönsson B. Visual analog domains for interpretation of back and leg pain intensity in patients operated for degenerative lumbar spine disorders. *Spine* 2001; 26 (21): 2375-80.
- Zanoli G, Strömqvist B, Jönsson B, Padua R, Romanini E. Pain in low-back pain. Problems in measuring outcomes in musculoskeletal disorders. *Acta Orthop Scand (Suppl 305)* 2002; 73: 54-7.
- Zanoli G, Strömqvist B, Jönsson B. Outcome assessment of spine surgery using the SF-36 questionnaire. Manuscript, quoted in Zanoli G Outcome assessment in lumbar spinal disorders *Acta Orthop (Suppl 318)* 2005; 76: 5-47.
- Zanoli G, Jönsson B, Strömqvist B. SF-36 scores in degenerative lumbar spine disorders. Analysis of prospective data from 451 patients. *Acta Orthop* 2006; 77: 298-306.