

From the Department of Orthopedics
Lund University Hospital
Lund, Sweden

Outcome assessment in lumbar spine surgery

Gustavo Zanoli



Thesis 2005

ACTA ORTHOPAEDICA SUPPLEMENTUM NO. 318, VOL. 76, 2005



Contact address

Gustavo Zanoli
Department of Orthopedics
Lund University Hospital
SE-22185 Lund, Sweden
tel: +46 46 172063
fax: +46 46 130732
email: gustavo.zanoli@ort.lu.se

Department of Orthopedics
Università di Ferrara
Clinica Ortopedica
Arcispedale S. Anna
44100 Ferrara, Italy
tel: +39 0532 237112
fax: +39 0532 209250
email: gzanoli@globeweb.org
www.globeweb.org

Cover illustrations: Nereo Quagliato (1939)
Sketch for a bronze statue of a young girl (1982)
Gouache on paper (32 x 48 cm)
Private collection

ISSN 1652-8220
ISBN 91-85439-47-9
Lund University, Faculty of Medicine Doctoral Dissertation Series 2005:41

Printed in Sweden
Wallin & Dalholm
2005

For Elisabetta
and Gabriele, Emma and Anna
our best outcomes
who need no assessment

List of papers

This thesis is based on the following papers, which in the text are referred to by their roman numerals:

- I. 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.
- II. Zanoli G, Nilsson LT, Strömqvist B. Reliability of the prospective data collection protocol of the Swedish Spine Register: test-retest analysis of 119 patients. Submitted.
- III. Zanoli 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.
- IV. Zanoli G, Strömqvist B, Jönsson B. SF-36 scores in degenerative lumbar spine disorders: analysis of prospective data from 451 patients. Submitted.
- V. Zanoli G, Strömqvist B, Jönsson B. SF-36 for outcomes assessment of spine surgery. Submitted.
- VI. Zanoli G, Hsu K, Zucherman J, Strömqvist B. SF-36 profiles before and one year after spinal stenosis surgery – a prospective comparison of two techniques in two nations. Submitted.

Contents

Thesis at a glance, 2

Abbreviations and definitions, 3

Introduction, 5

- Random reflections on (clinical) measurement, 5
- Pain in low-back pain and evidence-based practice, 6
- Philosophy and science, 7
- Steps to an ecology of clinical research: EBM and complexity in clinical practice, 8

Aims of the studies, 10

Methods, 11

- Literature search (Paper I), 11
- Data extraction (Paper I), 11
- Clinical Protocol (Papers II–VI), 11
- Patients (Papers II–VI), 12
- Diagnoses (Papers II–VI), 12
- Surgical procedures (Papers II–VI), 12
- Postoperative treatment (Papers II–VI), 13
- Outcomes (Papers II–VI), 13
- Statistical methods, 13
- Data entry, elaboration and analysis (Papers II–VI), 14

Overview of timeframe of studies, 15

Summary of papers, 16

- Paper I. Lessons learned searching for a HRQoL instrument to assess the results of treatment in persons with lumbar disorders, 16
- Paper II. Reliability of the prospective data collection protocol of the Swedish Spine Register: test-retest analysis of 119 patients, 16
- Paper III. Visual analog scales for interpretation of back and leg pain intensity in patients operated for degenerative lumbar spine disorders,

17

Paper IV. SF-36 scores in degenerative lumbar spine disorders: analysis of prospective data from 451 patients, 18

Paper V. SF-36 for outcomes assessment of spine surgery, 20

Paper VI. SF-36 profiles before and one year after spinal stenosis surgery – a prospective comparison of two techniques in two nations, 21

Reference tables, 23

Frequencies of diagnoses and procedures in the whole sample 1993–2003 (paper III, IV and V), 23

SF-36 reference values (paper IV and V), 23

Responsiveness (paper V), 26

Reliability (paper II), 27

MCIDs (paper V), 28

Discussion, 29

Conclusions, 36

Future perspectives, 37

Summary, 38

Acknowledgments, 39

References, 40

Summary in Swedish – Svensk sammanfattning, 43

Appendix, 44

Forms, 44

Literary and figurative references, 46

Final note, 47

Thesis at a glance

Question	Method	Answer (paper)
Has there been a uniform method of evaluating results of spine surgery over the past 30 years?	Websearch and systematic review of the literature	NO (I)
Is the protocol of patient-centered evaluation of the Swedish Lumbar Spine Register reliable?	Test-retest design with ICC and kappa statistics	YES (II)
Is VAS useful in describing different diagnostic profiles in persons operated on for lumbar spine disorders?	Mean-plots with 95% CI	YES (III)
Are prospective and retrospective pain assessments comparable?	Correlation with patient satisfaction	NO (III, V)
Is SF-36 useful in describing different diagnostic profiles in persons operated on for lumbar spine disorders?	Mean-plots with 95% CI	YES (IV)
Is SF-36 responsive to change in persons operated on for lumbar spine disorders?	Standardised Response Mean, Standardised Effect Size, paired T Statistics, Responsiveness Statistics	YES (V)
Is it possible to identify separate subgroups among persons operated on for lumbar spine disorders?	k-cluster analysis	YES (V)
Are preoperative clusters predictive of 12 months improvement?	Mann-Whitney of prospective and retrospective items	NO (V)
Is SF-36 useful in comparing outcome of spine surgery across nations?	Norm-based scoring	YES (VI)

Abbreviations and definitions

Diagnoses

HD	Lumbar disc herniation
CS	Central stenosis
DDD	Degenerative disc disease
SO	Spondylolisthesis
LS	Lateral stenosis

Outcomes

Domain – A subscore within a questionnaire, describing a specific area of interest within a larger questionnaire (e.g. subscales in SF36).

HRQoL – Health-Related Quality of Life.

Likert scale – A rating scale with a limited number of predetermined possible answers or states, categorised in an ordinal rank.

Outcome – The end result of a clinical intervention.

Outcome research – Study of the measurement properties of outcome indicators, comparing different methods and instruments.

Outcome assessment – The clinical activity of collecting outcomes in a standardised fashion to be able to interpret personal results.

Questionnaire – A set of questions submitted to persons receiving health care in order to share their view on the treatment received and perform an outcome evaluation, usually patient-administered. Tool, instrument survey or tool are often used as synonyms.

Patient-centered – Referred to outcome instruments, it indicates a tool (usually a questionnaire) which is completed directly by the patient with no interaction from the caregiver.

Objective – Referred to outcome instruments, it indicates a dimension which is measured directly by the assessor without subjective interpolation.

(Disease)specific – Referred to outcome instruments, it indicates a tool (usually a questionnaire) which is focusing on one or more aspects (Function, impairment etc) of a specific health problem that is affecting the person.

Generic – Referred to outcome instruments, it indicates a tool (usually a questionnaire) which is applicable to a general population, regardless

of the specific health problem that is affecting the person

RSA – Radiostereometric analysis

SF-36 – Short Form 36-item questionnaire derived from the Medical Outcome Study:

PF SF-36 – Physical Function

RP SF-36 – Role Physical

BP SF-36 – Bodily Pain

GH SF-36 – General Health

VT SF-36 – Vitality

SF SF-36 – Social Function

RE SF-36 – Role Emotional

MH SF-36 – Mental Health

VAS – Visual analogue scale, usually a 10 cm horizontal line where a dimension can be marked. Usually quantified as 0–10 or 0–100.

Statistics

Cluster analysis – An exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise. Given the above, cluster analysis can be used to discover structures in data without providing an explanation/interpretation. In other words, cluster analysis simply discovers structures in data without explaining why they exist. (www.statsoft.com/textbook/stcluan.html)

ICC – two-way random model Intraclass Correlation Coefficient.

K – weighted kappa statistics, provides a measure of agreement between two observations.

MCIDs – Minimal clinically important differences.

95% CI – 95% confidence interval

Clinical research

RCTs – Randomised Controlled Trials

SRs – Systematic reviews: a type of scientific study which aims at reviewing all existing publications on a certain subject.

Meta-analysis – A statistical tool that can be applied to results of a SR to summarise them and

reach conclusions regarding the research question.

Cochrane Reviews – Systematic reviews of RCTs that try to answer therapeutical questions.

Register – A prospective observational study, involving clinical data collection. National reg-

isters for knee and hip arthroplasty in Sweden and Norway have become a reference point for scientific and clinical information. Similar aims have led to the promotion of the National Lumbar Spine Register in Sweden.

Introduction

Those diseases which medicines do not cure, iron cures; those which iron cannot cure, fire cures; and those which fire cannot cure, are to be reckoned wholly incurable (Hippocrates, Aphorisms, around 400 BC). The last of the famous Hippocratic aphorisms is quoted by Girolamo Fabrizio d'Aquapendente (1533–1619) in his surgical textbook as a testimony in favour of the power of surgery over medicine. And this respect – when not deference – toward the great surgeon, with “gifted” hands, has remained throughout the centuries until our time, even in small linguistic nuances: for example, patients still *submit* to surgery. The great technological improvements in pain and infection control of the last century have enabled surgeons to achieve in a regular and predictable mode what before used to be only a lucky possibility. However, this has also contributed in placing higher expectations on the results of surgery, both from surgeons and patients and has produced in some cases an excess in the indications for surgery which might be both ethically incorrect in the single case as well as economically unsustainable for the society. In today's hard competition for resources, the (cost-)effectiveness of many spinal surgeries has been questioned (Gibson et al. 1999, Keller et al. 1999, Deyo et al. 2004). A great variability in practice across regions and nations (Jansson et al. 2004), together with the absence of well designed RCTs in the earlier guidelines can be considered as the basis of this “inquiry”. Although new RCTs are coming which address the relevant questions (Fritzell et al. 2001, Brox et al. 2003, Fairbank et al. 2004), their interpretation has not always been straightforward, and this difficulty is reflected in the recent publication of the European guidelines on chronic low-back pain (www.backpaineurope.org). Spinal surgery has entered the era of Evidence-Based Medicine (EBM) and surgeons accept to play by the rules in the scientific community, but many aspects of research methodology are still not completely standardised and, at the same time, the complex picture of the people who suffer from low-back pain seems to elude all standardisation

efforts. The *art of measuring* (Roos 1997) is still necessary for the humble clinical researcher who is walking on *jagged bits of bottle* just as much as the *wounded surgeon* needs to learn the difficult art of equilibrium between the *pure too-little* and the *empty too-much* when applying scientific knowledge to our handicraft specialty (see Appendix).

Random reflections on (clinical) measurement

To assess outcomes after lumbar spine surgery we need measures, and measures are quantified by numbers. Obviously, not all numbers serve as outcome evaluation: for example the progressive chart identification number of a patient is only useful to trace the record of this person. Other numbers actually mean something (days of sickleave, VAS) but need methodological work before they can be considered good measures, arithmetically and statistically useful. But even when we have a good measure we cannot assume that we are measuring any True value. In physics, it has been known for a long time the measurement *alters* the object, quantum theory (e.g. Heisenberg uncertainty principle) has gone even further by saying that measurement *creates* the object, at least in the form which is measured. More recently, philosophers like Edgar Morin have reflected on the issue of complexity and suggested that it is necessary *to face this complex problem in which the knowing subject becomes object of its knowledge while remaining subject* (Morin 1986). It may seem impossible for a clinical researcher to deal with the multifaceted problems of outcomes assessment, but it is necessary today as it was a century ago in Boston where Ernest Amory Codman applied his “End Result Idea” not only to scientific papers and communications, but to the everyday evaluation of “the product of an Hospital”. The main difference today is that we no longer can believe in a simplistic possibility to *master reality* (classifying errors and outcomes and providing deterministic management solutions): in

fact it has been noted (Hägg 2002) that this “almost century-old vision of an evidence-based medicine and patient-centred hospital administration is even to this day, at least in Sweden, nowhere completely enterprised”. What we can learn from other scientific disciplines and the philosophic reflection is that we need to *have a dialogue with reality* (Browaeys 2004), rather than believing in the possibility to quantify everything. For examples we have to accept and expect distortions such as the Hawthorne effect: an experimental effect in the direction expected but not for the reason expected; i.e. a significant positive effect that turns out to have no causal basis in the theoretical motivation for the intervention, but is apparently due to the effect on the participants of knowing themselves to be studied in connection with the outcomes measured (Draper 2005). The Hawthorne effect is somehow comparable to the Heisenberg principle in physics, with the difference that it is not as constant and predictable. Trying to understand these concepts, it might be useful to add to the rational arguments some examples from figurative arts, where a similar evolution can be observed. In the same years Codman was formulating his ideas, Bonnard was painting by heart impressionistic gardens that are still quite recognisable (Figure 15), whereas the younger Klee’s remembrance of a garden (Figure 16) challenges the traditional iconography. Which of the two is closer to which Truth?. Few years later, Magritte started his famous “*Ceci n’est pas une pipe*” series which he continued until the year before his death (Figure 17). Maybe sometimes, as clinical researchers dealing with outcome assessment, we should remember that “this is not an outcome” but only one of the possible representations of the human condition we want to measure.

Pain in low-back pain and evidence-based practice

People seeking help from the lumbar surgeon usually have pain in the back or in the leg(s) or both. Sometimes is difficult to differentiate among them, even in the bible the injury suffered by Jacob when fighting against the Angel (Gen 32, 23–33) is described as a hip dislocation but is often quoted as an example of sciatica. To increase confusion,

the keyword *low-back pain* has become a standard indication for a broad group of conditions in the biomedical literature, as if it was a diagnosis rather than a simple descriptive symptom. It is thus unavoidable for the clinical researcher to deal with the issue of pain measurement in lumbar spine surgery.

The most straightforward way to measure pain is probably simply asking the patient to quantify it on visual analog scale (VAS) (Huskisson 1974) or on a numerical radical scale (NRS) (Jensen et al. 1986). Both approaches have been used more than once in LBP (Evans and Kagan 1986, Schofferman 1999). Another even more commonly used approach is to submit to the patient an ordinal scale of attributes that describe pain (for example no pain, mild pain, some pain, a lot of pain, worst pain), and goes under the name of Verbal Rating Scale (VRS): we should probably speak of verbal rating scales, as selection of words and number of possible choices in the ordinal scale vary from study to study. In a methodological perspective these scales can be considered as patient-oriented measures, even though it is likely that before the dissemination of knowledge about self-administered questionnaires, VAS, NRS and VRS were actually collected by physicians in many studies.

From a formal point of view, these are single items, as opposed to the more compound multi-items questionnaires, which have been developed recently. Out of 15 specific pain questionnaires found in our review (Paper I), 5 were personal sets of question, designed for a specific study, 5 were proposed questionnaires, which did not undergo further evaluation, and 5 were validated questionnaires, which underwent psychometric or clinimetric testing also in LBP settings. Reliability (the ability of a given instrument to give reproducible measures of the measured entity), validity (the ability of a given instrument to give “true” measures of the measured entity), and responsiveness (the ability of a given instrument to detect a known change in the measured entity), are usually studied in these undertakings, which very often provide also useful normative data to enable correlation with different population samples and power calculation. For these reasons, and to encourage the use of very few “standards” which allow comparisons between studies, validated questionnaires are preferable.

If we concentrate on the substance of measurement, we can assume that, unless otherwise specified, VAS, and NRS are intended to be measures of pain intensity. It is important, when trying to extract evidence from a study, to check if this was actually the case. For the same reason it would be important that authors specify what they mean by VAS measurement. VRSs usually clarify this aspect with the choice of the terms included in the administered scale. Measuring pain intensity seems to be by far the most common approach to the quantification of pain, but is not the only one. Pain-related disability can be connected with pain intensity to assess pain severity. This happens for example in the SF-36 bodily pain domain, which combines a question on pain intensity with one on interference with activities.

More recently patient-specific questionnaires have been introduced, which include self-rated importance of symptoms or items to calibrate the results (Chatman et al. 1997). A different approach to quantification of pain, can be the assessment of number of days (hours) with pain. This can be done either retrospectively or in a prospective manner (diary). These methods can be important in the case of chronic and/or recurrent pain, two situations which are quite different from the acute perspective, and for which we still lack accepted definitions and dedicated instruments. In fact, most of the measures in use have been subsequently adapted from the acute setting where they were originally validated, and it has been shown that they possess different measurement properties in the various situations (see below).

Other aspects of the experience of pain can be taken into account, such as pain affect, which may be defined as the ensemble of emotions and feelings that accompany the sensation of pain itself. For a more in-depth review of the evidence on this subject, see (Von Korff et al. 2000).

As stated elsewhere (Zanoli 2002), after ten years of EBM many of the dreams of Archie Cochrane – the visionary epidemiologist under whose name the international collaboration for EBM has been founded – have come true: databases of RCTs and systematic reviews, secondary publications and many other sources of evidence-based publications are now available to the practicing clinician and to medical students, and in many fields of medi-

cine the ideal of an evidence-based clinical choice is more realistic than ever. This does not mean that all problems will be solved simply applying the method to all possible relevant clinical questions. For example, an important aspect of pain in chronic diseases or psychosocial conditions such as low-back pain is the possibility to understand and accept the symptoms or the ability to develop coping strategies. This is not only a matter of personal attitudes, the information that people receive from their health care providers and the different clinical pathways they follow might influence their acceptance of pain. It is true however, that even with comparable intensity of pain the same therapeutic program might not be acceptable for all: some people would try just anything before undergoing surgery, others believe in surgery as a possible *deus ex machina* shortcut. It is possible that people who seek help from the surgeon are those who fail to cope with their symptoms, and are therefore *suited* for surgery, whereas the ones who are *suited* for physiotherapy are those who like to be empowered of their own health. On the other hand, it might be that patients selected for surgery or multidisciplinary approach are not *truly* different but rather captured in those two different states just like electrons can be measured at the same time as waves or particles in physics. This is however a good example of evidence which can be interpreted differently by different people in different clinical and cultural contexts. The idea of a True Evidence which deterministically informs all clinical decision is only in the view of simple-minded EBM enthusiasts.

Philosophy and science

After using examples from other scientific disciplines and art as a source of inspiration to explain some of the concepts, which are in the background of this thesis, a tentative summary from some discontinuous and personal reading of the present epistemological discussion can serve as a basis to defend the reason for this work.

Epistemology is the study of the scientific knowledge from a critical point of view, analyzing the principles, hypotheses and results of the sciences to determine their value. The revolution

caused by the quantum mechanics in physics has not only greatly influenced our everyday life (personal computers, laser, automatic doors and speed detectors are outcomes of its application) but it has also attracted many philosophers to reflect on the (r)evolution of scientific theories, and because of the peculiar implications of this theory on philosophical issues like reality, objectivity and causality, it has also confused the borders between epistemology, theory of knowledge, philosophy of science and philosophy itself. Classic works by Karl R. Popper (1902–1994) have clarified the typical process of scientific progress, starting from previous *conjectural* knowledge to confute it through hypothesis testing and never accepting a confirmation as a final Truth but *conclusionally refuting* it whenever it fails at the test of experiments (Logik der Forschung 1935, Conjectures and Refutations 1963). A similar mechanism is at the basis of Thomas S. Kuhn's (1922–1996) definitions of *paradigm shifts*, as a response to crisis in the stable route of *normal science*, i.e. when enough evidence is accumulated against the mainstream theory to throw it away and let the new paradigm try to solve the problems which remained after abandoning the old one (The structure of Scientific Revolutions 1962). Gregory Bateson (1904–1980) has explored the relationship between Mind and Matter, Organisms and Environment, Ecology and Tautology considering epistemology both from a scientific as well as a philosophic point of view. The complex interactions of the many parts that constitute a mind can be studied – according to Bateson – in the same way ecology explores relationships between organism and environment (Steps to an Ecology of Mind 1973, Mind and Nature 1980). He opposes ecology and tautology in a way which somehow resembles the conflict between normal science and paradigm shift proposed by Kuhn (and, with a different attitude towards normal science, by Popper). The complex issues of applying scientific knowledge to clinical decision making can benefit of the work of John Mackie (1917–1981) introducing the concept of an *INUS condition*: “an insufficient but necessary part of a condition which is itself unnecessary but sufficient for the result” (Causes and Conditions, 1965). Similar conditions can therefore lead to different results. Present philosophical issues raised by the theory of complexity,

a multidisciplinary science, that takes into consideration elements of very different disciplines, such as systems theory, cybernetics, meteorology, chaos theory, artificial intelligence (A.I.), artificial life, cognitive sciences, computer science, ecology, economy, evolution studies, genetics, games theory, immunology, linguistics, philosophy, social sciences, management.

Edgar Morin (1921–) – one of the heralds of this new paradigm – invites us to accept the contemporary presence of contrasting concepts such as order and disorder that do not exclude themselves, but coexist and find a dynamic equilibrium.

Steps to an ecology of clinical research: EBM and complexity in clinical practice

This caricatural reduction of deep philosophical research work is in our intention necessary to justify the logic that has informed the work behind this thesis and the (sometimes) invisible ties that bind together the papers presented. In our interpretation, RCTs can be considered the industrial revolution in clinical research, sustained by the positivistic view that any answer is possible, it's just a matter of time (and money): sir Francis Bacon, whose utopistic views are – according to Popper – still the basis of the enduring “religion of Science”, was similarly optimistically convinced that in a couple of years with the proper funding he could explain all the book of Nature in the light of his *new science*. Cochrane reviews and meta-analyses are indisputably the highest achievement of this approach, which has certainly helped clinical medicine out of the desert of authoritarian experience and the low waters of physiopathological reasoning. But EBM as a new paradigm can still go a longer way, if it accepts to deal with the issues of complexity, responding to the crisis raised by clinical dilemmas by incorporating the awareness of the effects of the environment and on the environment.

In a systematic review of RCTs on hip arthroplasties recently completed for the Italian NIH (Mele 2004), the conclusion suggests the need to expand the review to (high quality) observational data in order to be able to answer the relevant questions in this field. RCTs are not only too expensive

and too slow for this technology, they might also be too far away from real clinical practice. So the future of clinical research could be to incorporate other types of evidence, such as the ones which come from a rigorous and reliable outcome assessment: this should be possibly achieved with less resources (funding) and still be closer to real life (i.e. to the problems of persons who seek healthcare). This *ecological* approach is the basis of the papers which are collated in the present thesis: there was no formal hypothesis testing at the basis of this work, instead the strongest steering idea was that of *recycling* otherwise lost complex data. Not only to avoid a possible publication bias (which is considered one of the biggest threats to EBM as a whole) but also not to waste the small bits of truth that this (partially incomplete) data contained. Limitations of such retrospective *a posteriori* design will be addressed in the discussion (as many times before in the responses to peer reviewers): most of the times it was quite difficult to make the usefulness of this methodology clear. Paradoxically, even

in the EBM era, it is still more easy to “sell” (i.e. to publish) meaningless case series presenting a new technique with no sound methodological approach. We also many times have been discouraged ourselves by the complexity of the clinical picture that emerged from outcome assessment and its translation into clinical practice. In a time of globalisation, it is easier to digest low-quality *fast-food* information, rather than sitting down and elaborate on our own clinical results. Founded in reaction to the opening of the first McDonald in Rome, the *slowfood* movement (www.slowfood.com) has succeeded in bringing the issues of taste, methodological quality and ecological sustainability to the world’s attention. As for most biomedical research, even the information contained in the manuscripts presented will be quickly digested and thrown away. But maybe the methodological contribution can be recycled in other ways and help some of the readers in “preparing” their own “food”, assessing their own outcomes. At least as long as a new paradigm arises.

Aims of the studies

The overall aim of this thesis was to increase understanding of some aspects of outcome measurement in patients operated on for lumbar spine problems using existing prospective data available at the spine section of the Orthopedics Department of the University of Lund, and, moreover, to explore potentials and weaknesses in the methodology of retrospective analysis of prospectively collected observational data.

Specific aims were:

Paper I – To propose a systematic approach to the search of evidence in Medline and Internet. Information provided in the article should facilitate the choice of a health-related quality of life (HRQoL) instrument and future research in this field. Moreover the awareness of the existence of many established instruments could discourage potential “wheel re-inventors”, thus diminishing the number of unnecessary new questionnaires.

Paper II – To assess test-retest reliability of the prospective data collection protocol in use by the Swedish Spine Register and raise the issue of quality of data in large observational studies.

Paper III – To illustrate the use of recording VAS for pain intensity in patients operated on for lumbar spine problems, in particular to describe differences between individuals or groups of people and evaluate change over time. Compa-

rability of different methods of prospective and retrospective pain assessment was also investigated.

Paper IV – To determine the preoperative SF-36 scores in patients scheduled for lumbar spine surgery as a source of reference for future studies and to illustrate the use of SF-36 in describing differences between groups of people with a similar diagnosis.

Paper V – To describe the use of SF-36 in assessing 1-year outcomes of patients operated on for lumbar spine disorders, mainly providing four types of information: 1) postoperative profiles for SF-36 and mean improvements in lumbar spine patients for comparison between existing and future studies, 2) reference values for improvement to serve as reference when MCIDs values are needed (e.g. sample size calculations), 3) responsiveness of VAS and SF-36 physical domains and estimates of effect sizes, and 4) identification of clusters from the preoperative scores and analysis of their postoperative profiles.

Paper VI – To investigate the possibility of using SF-36 when comparing two groups of patients with spinal stenosis who underwent different surgical procedures in two separated prospective studies in two different countries.

Methods

Literature search (Paper I)

A systematic Medline search was performed using following search strategies:

1. Low Back Pain [MESH] AND Outcome Assessment (Health Care) [MESH] AND Questionnaires [MESH]
2. (outcome OR outcomes) AND (survey OR surveys OR questionnaire OR questionnaires) AND low back pain
3. absence of language limitation.

The search was run a first time at the beginning of 1999 and subsequently updated until the end of 2000 (just before printing the final version).

Similar searches were also performed with common Internet search engines and specific websites: American Association of Orthopaedic Surgeons (www.aaos.org), MAPI (www.mapi-research-inst.com), Medical Outcomes Survey (www.outcomes-trust.org), Quality of Life Assessment in Medicine (www.qlmed.org), International Society of Quality of Life Research (www.isoqol.org), Health Outcomes Institute (www.hoi-stratishealth.org), and the Cochrane Library (www.update-software.com/clibhome/clibip.htm).

Data extraction (Paper I)

Abstracts of all relevant articles were read in the search for original instruments or adaptations. The paper was included in the list if authors stated that they had used some type of patient-centred questionnaire in the evaluation of results. Questionnaires developed for epidemiologic or occupational purposes were excluded, outcome scores that included patient-centered items were also included. Criteria for the evaluation of retrieved instruments were validity, reliability, responsiveness and availability in different languages.

Data were extracted from the papers to an Excel spreadsheet, using following headings and categories to analyse the material: **Author**, **Full Reference**, **Journal**, **Year**: year of first publication on

the original instrument. **Name**: When applicable, acronym of the instrument. **Type**: *Gen*: generic instrument, *Spec*: specific (functional) instrument, *Pain*: pain measurement, *Psy*: psychological assessment, *Sat*: satisfaction instrument. **Validity**: *Per*: personal set of questions, *Pro*: proposed questionnaire (some validation), *Sco*: score of different parameters, even non-subjective, *Val*: validated questionnaire. **Specialty**: *GP*: general practice, *NC*: non-conventional, *NE*: neurological sciences, *OT*: orthopaedics and traumatology, *PA*, pain/anaesthesiology, *PH*: public health, *PSY*: mental disciplines, *RE*: physiotherapy, *RH*: rheumatology. **Source**: *M*: Medline, *P*: previous knowledge; *R*: referenced. **Reference in LBP**: Study in which the instrument was applied to LBP patients (only for instruments not originally designed for LBP).

Clinical protocol (Papers II–VI)

Encouraged by prosperous registries for hip and knee arthroplasty, the idea to develop a Swedish registry for lumbar spine surgery was born and developed during a meeting on degenerative lumbar spine disorders in Lund 1992. This state-of-the-art conference, among other issues, discussed the layout of a protocol for lumbar spine surgery that was published in 1993 (Strömquist and Jönsson 1993). The protocol – which was adopted immediately at the spine section of the Orthopaedics Department of the University of Lund – consisted of preoperative subjective data and objective evaluation completed by the surgeon, surgical data and a follow-up form completed 4, 12 and 24 months postoperatively (see Appendix). The protocol was approved by Datainspektionen, later, after change of regulations, according to the Personuppgiftslagen (PUL). It was supported by the Swedish National Board of Health and Welfare from 1993 and also adopted by the Swedish Society for Spinal Surgery but it did not obtain a nationwide dissemination until 1998 when the register was transferred to the Society and administrative and

Table 1. General objective criteria for the five diagnostic groups

Diagnosis	Criteria for surgical indication
Lumbar disc herniation	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. One of the two 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. Radiography and MRI or CT confirmation.
Lateral stenosis	Radiating pain in a single dermatome and/or neuromuscular deficits at the same level and/or mechanical LBP. MRI or CT confirmation.

content changes were introduced (Strömqvist et al. 2001). The new protocol excluded objective examination and was made entirely patient-based with 13 questions (see Appendix) and the introduction of SF-36 as a HRQoL measure. The patients independently were asked to complete the preoperative questionnaires on the day of admission for surgery and postoperatively after 1-year during a follow-up visit or, if no visit was planned, by mail using a prepaid envelope. Further changes were introduced in 2003 which did not affect the material presented in this thesis. More information and downloadable forms are available at www.4s.nu.

Patients (Papers II–VI)

Since 1993 all persons undergoing elective lumbar surgery at the spine section of the Orthopedics Department of the University of Lund were included in the prospective registration protocol either in its first version (1993–1997), which was the source of data for paper III, or in its revised version after 1998, which was the source of data for papers II, IV, V, and VI. Tumours and fracture cases are excluded from the register.

In paper II, 63 persons were asked to fill in the preoperative questionnaire twice between January 2000 and March 2003, during the same period 59 completed the postoperative questionnaire twice.

In paper VI, data from 90 persons who had undergone elective lumbar surgery for spinal stenosis was extracted serving as an age- and sex matched comparison to patients operated on in a US prospective FDA study with a new technique.

The register can be considered a prospective

observational (large) cohort study design, for this reason it is not incompatible with any other concurrent experimental study performed on the same person. Some of the persons registered in the protocol also participated in one or more experimental studies (RSA cohort studies, RCTs) which were performed at the spine section of the Orthopedics Department of the University of Lund during this 10 years period. They were not excluded from our analysis and were classified according to their respective diagnostic group and intervention type.

Diagnoses (Papers II–VI)

Diagnoses were entered by the individual surgeons from a multiple-choice list but no predetermined set of criteria was used. The criteria presented in Table 1 were set a posteriori to help the interpretation of the data, and should not be regarded as strict and cogent definitions. The protocol also allowed to enter *Other* or *Combination* in this field, if the surgeon did not find the five categories described properly the condition. However, this happened in a very small number of cases (5.5%).

Surgical procedures (Papers II–VI)

Surgical procedures presented in Table 2 were entered by the individual surgeons from the list. The protocol also allowed entering *Other* or *Combination* in this field, if the surgeon did not find the predetermined categories described properly the condition. This, however, occurred in a very small number of cases (1.7%).

Table 2. Elective surgical procedures performed

Surgical procedure	Specification
Conventional discectomy	Onik
Microscopic discectomy	
Percutaneous discectomy	
Decompression	Bak
Instrumented anterior fusion	
Uninstrumented posterior fusion	
Instrumented posterior fusion	VSP, CCD, Diapason,....
Decompression + uninstrumented posterior fusion	VSP, CCD, Diapason,....
Decompression + instrumented posterior fusion	

Postoperative treatment (Papers II–VI)

The standard protocol for postoperative treatment at the spine section of the Orthopedics Department of the University of Lund consists of physiotherapy at 6 weeks after decompressive surgery and at 4 months when fusion is included in the procedure. However, this information was never included in the registration protocol, nor was it controlled in any way that the patients actually complied with the prescription. Although concomitant treatment may play a role in determining the outcome of surgery, this was felt not be a limitation of the present thesis, since it is dealing mainly with the methodology of outcome assessment.

Outcomes (Papers II–VI)

Initial information at baseline (the day before surgery) included age, sex, smoking habits, duration of preoperative back and leg pain in months, duration of preoperative sickleave in months, number of previous operations.

Patient's working status was recorded on a 6-point Likert Scale (heavy work, medium heavy work, light work, disability pension, unemployed, retired).

The utilisation of several diagnostic techniques (CT, MRI, myelography, MmyeloCT, nerve root injections) during the diagnostic workout was recorded as a binary variable (yes/no).

Pre- and postoperatively VAS scores were obtained in standard fashion, i.e. by measuring the distance in mm from the origin of an horizontal line (total length 100 mm) and the point indicated by the patient as representing his/her level of pain during the preceding week. Zero represented "no pain at all" and 100 "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 to 1 km, more than 1 km). Swedish version (Sullivan et al. 1994) of the SF-36 survey (Ware and Sherbourne 1992) was added to the protocol in 1998.

Postoperatively, change in leg and back pain, respectively, was recorded on a 5-point Likert Scale (Painfree, Much Better, Somewhat Better, Unchanged, Worse) as compared to preoperative status. 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 disorder, sick leave because of other problems, unemployed, retired). Patient satisfaction was recorded on a 3-point Likert Scale (satisfied, uncertain, dissatisfied).

Individual SF-36 scores were obtained for each patient pre- and post-operatively, and the change in scores between the two time points was calculated. In addition, individual differential values from normative national age- and sex-matched scores were determined (Ware et al. 1993, Sullivan et al. 1994, Ware et al. 2000). Differential values (postoperative values *minus* preoperative values) were calculated also for VAS scores.

Comparison between preoperative and postoperative responses to the walking distance item allowed for an *a posteriori* classification into 4 categories (worsened, unchanged, improved and no limitations, for those who remained stable over 1 km walking distance pre- and postoperatively), which was used only in paper II.

Statistical methods

Since very often continuous data for which no normal distribution could be demonstrated (like age) were analysed together with ordinal and nominal scales, non-parametric methods were utilised

Run #	Social sec. #	Name	Delete	Pre	SF1	Op	4m	12m	SF2	24m	Reop	mo	Dead	Valid
LUN 1	000000-0000	Doe, John	☐	OK	OK	OK	OK	OK	call	call			OK	
LUN 2	000000-0000	Doe, Jane	☐	OK	OK	OK	OK	unf	reop	reop	OK	18	OK	

Figure 1. Computer application. Starting page with flip-top head.

for descriptive purposes (median), for correlation analyses (Spearman Rho) and for significance testing, e.g. investigating the presence of a selection bias in demographic and preoperative characteristics among persons who returned/did not return at different follow-ups (Mann-Whitney) or comparing pre and postoperative values on the same sample (Wilcoxon). In paper III, 95% confidence intervals for median values were calculated according to Bland (Conover 1980).

In descriptive analyses, SF-36 scores were analysed with parametric methods (mean and 95% CI), in accordance with the standard way of reporting in the literature.

In paper II 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 answer levels (7 for RP and 5 for RE), both ICCs and kappas were calculated.

In paper V, “K-means” cluster analysis was based on each patient’s preoperative norm-based scores in the eight scales of the SF-36. 2, 3 and 4 clusters discrimination was performed and results plotted to verify discriminative ability. Responsiveness of relevant variables was investigated using 4 different methods as suggested by Walters and Brazier (2003): Standardised Response Mean, Standardised Effect Size, paired T Statistics and Responsiveness Statistics.

Data entry, elaboration and analysis (Papers II–VI)

Patients and surgeons always completed surveys in paper form. Questionnaires were collected centrally and entered into a personalised Filemaker Pro database (Knutson 1993–2002, Figure 1) by the same person. When appropriate, questionnaires were scored directly by the application which implemented the original scoring algorithms.

Subsequently data was extracted to an Excel file, which was used to calculate change in scores and differences from norm and to format the file for the statistical package.

Statistical analysis was performed by using dedicated software (SPSS for Windows 11.5, SPSS Inc). MedCalc 7.6.0.0 (www.medcalc.be) was used for weighted kappa calculation in paper II.

Overview of timeframe of studies

Paper I – Websearch and systematic review of the literature performed in 2000. No patients involved

Paper II – 122 patients undergoing elective lumbar surgery at the Department of Orthopedics of the University of Lund from 2000 to 2003 who completed the questionnaire of the Swedish Lumbar Spine Register twice (test-retest).

Paper III – 755 patients undergoing elective lumbar surgery at the Department of Orthopedics of the University of Lund from 1993 to 1998 who completed a preoperative VAS for intensity of pain were included in the study.

Paper IV – 451 patients undergoing elective lumbar surgery at the Department of Orthopedics of the University of Lund from 1998 to 2002 who completed the preoperative questionnaire

of the Swedish Lumbar Spine Registry were included in the study.

Paper V – 351 patients undergoing elective lumbar surgery at the Department of Orthopedics of the University of Lund from 1998 to 2002 who completed the pre- and postoperative questionnaire of the Swedish Lumbar Spine Registry were included in the study.

Paper VI – 90 patients undergoing elective lumbar surgery for spinal stenosis at the Department of Orthopedics of the University of Lund from 1998 to 2001 who completed the pre- and postoperative questionnaire of the Swedish Lumbar Spine Registry were included in the study (and served as age- and sex matched controls to 90 patients operated on in a US prospective FDA study with a new technique).

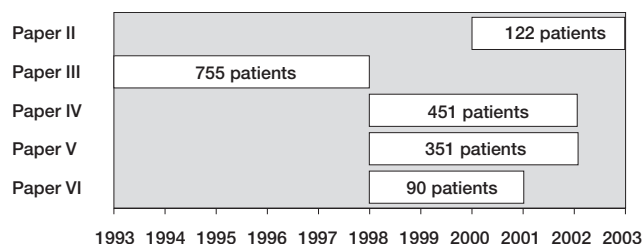


Figure 2. Overview of timeframe of studies, 1993–2003

Summary of papers

Paper I. Lessons learned searching for a HRQoL instrument to assess the results of treatment in persons with lumbar disorders

Study design. Websearch and systematic review of the literature

Background. Choosing the appropriate outcomes to measure in everyday practice is a typical clinical decision which should be based on the best evidence in the published literature or in other sources of retrievable information. The issue of standardizing outcome evaluation in assessing the results of treatment in persons with lumbar disorders has been raised before, but only “expert” solutions have been provided.

Objective. To propose a systematic approach to Medline and Internet information that could identify evidence to take into account in the choice of a health-related quality of life (HRQoL) instrument.

Methods. A systematic Medline search was performed using following search strategies:

1. Low back pain [MESH] AND Outcome assessment (Health Care) [MESH] AND Questionnaires [MESH].
2. (outcome OR outcomes) AND (survey OR surveys OR questionnaire OR questionnaires) AND low back pain.

Similar searches were also performed with common Internet search engines and specific websites.

Abstracts of all relevant articles were read in the search for original instruments or adaptations. The article was included in the list if authors stated that they had used some type of patient-centred questionnaire in the evaluation of results. Questionnaires developed for epidemiologic or occupational purposes were excluded, outcome scores that included patient-centered items were also included. Criteria for the evaluation of retrieved instruments were validity, reliability, responsiveness and availability in different languages.

Results. More than 90 different instruments have been used in the last 30 years of 20th cen-

tury in approximately 500 published clinical evaluation articles on LBP in the medical literature. The number is more likely to be underestimated, because it is sometimes difficult to identify an individual set of questions used in clinical reports. Very few instruments have been used in more than one study or by different authors without being modified. In 1999, 30 papers have used specific HRQoL instruments in clinical evaluation of LBP: among those, there were 13 researchers who introduced new questionnaires or used non-standardized ones, 12 who used pre-existing ones, and 5 who tried to compare existing outcome measures.

Data on cultural adaptation in different languages are often not published and are not so easily accessible. Only a minority of the questionnaires seems to have been officially translated and validated for use in languages other than English, even though these data could be underestimated.

Conclusion. Despite the great amount of work and evidence accumulated, basic methodological issues of outcomes assessment in spinal diseases remain unsettled. This is reflected in a lack of standardization in the choice of outcome measures to evaluate in clinical studies. The number of proposed outcomes, outcome scores, outcome instruments is incredibly high: no new HRQoL instruments specific for LBP are needed. Instead, effort should be directed toward comparative studies and systematic reviews of the evidence available. Standardization of outcome evaluation should be a major task for international spine societies.

Paper II. Reliability of the prospective data collection protocol of the Swedish Spine Register: test-retest analysis of 119 patients

Study design. prospective cohort study with repeated measurement (test-retest).

Background. The Swedish Lumbar Spine Register is collecting patient-based data since 1998, and more than 80% of all spinal units in Sweden

are now including patients. In a few years it will produce useful clinical information, as arthroplasty registers have, but to allow the necessary interpretation, reliability of the protocol must be tested.

Objective. Assess test-retest reliability of the prospective data collection protocol in use by the Swedish Spine Register.

Methods. Between November 2000 and March 2003 a sample of 122 patients (63 males and 59 females, median age 53, range 22–84) was asked to fill in the questionnaire twice: 63 received it at the time of last preoperative visit (to compare results with the preoperative questionnaire) and were asked to fill in the questionnaire at home and bring it back at time of hospital admission, 59 received a second copy at the time of the 1-year follow-up visit (to compare results with the 1-year follow-up) and were asked to fill in the questionnaire at home after one week and then return it to the department by mail. Patients could only belong to one of the two groups. In the same period further 302 patients were registered: a Mann-Whitney test was performed to ensure that the sample did not show significant differences from the total material as regards demographics and preoperative variables.

Results. Patients answered the two questionnaires after different time spans (range 0–235 days), so separate reliability analysis was performed in relation to the time interval. 0–2 days produced a significant memory effect, after 3 weeks the reliability seemed to drop in the preoperative group, whereas postoperatively results seemed reproducible up to 9 weeks.

In the “worst case scenario”, i.e. analyzing the total material irrespective of time elapsed between 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 according to kappa values. Most other values were above 0.70, and for non-SF variables they ranged 0.79–0.89. Kappa values for the other ordinal outcome questions were high (0.74–0.91).

Conclusion. Researchers could use present data to guide the choice and interpretation of outcomes in their specific clinical setting. The protocol studied can reliably detect postoperative improvements between large groups of patients such as in a Register.

Paper III. Visual analog scales for interpretation of back and leg pain intensity in patients operated for degenerative lumbar spine disorders

Study design. Retrospective analysis of prospectively collected cohort data

Background. There is no consensus regarding pain outcomes assessment in spine patients. Pain intensity, recorded on a VAS, is one of the most used measures. Still, many aspects of its interpretation are still debated or unclear.

Objective. To describe the use of recording VAS for pain intensity in patients operated on for lumbar spine problems.

Methods. A total of 755 consecutive patients, mean age 50 (15–86) years, operated from 1993 to 1998 were included in the study; there were 420 males and 335 females. Diagnoses included lumbar disc herniation (45%), central stenosis (19%), lateral stenosis (14%), isthmic spondylolisthesis (9%), and degenerative disc disease (9%). Local pain, radiating pain, analgesic intake, and walking ability were recorded before surgery and at 4 and 12 months after surgery. The patients' opinions regarding change in pain and satisfaction with the result were assessed separately. Correlation among variables reflecting perceived pain was sought.

Results. A stable differentiation in 3 groups for back pain and in two for leg pain was obtained comparing preoperative VAS means for different diagnoses, as shown in Figure 3 and 4. Differences were significant between HD (minor back pain, pronounced leg pain) CS (moderate back pain, pronounced leg pain) DDD (pronounced back pain, minor leg pain) and SO (moderate back pain, minor leg pain). Among other preoperative variables studied, age showed a similar discriminating performance. Significant but moderate correlation between different types of pain outcomes and with patient satisfaction was present in all cases (Table 3).

Conclusion. Measuring pain intensity on a VAS is a useful tool in describing patients scheduled for lumbar spine surgery. Prospective and retrospective methods of pain assessment are not directly comparable. In the search for a standard in the evaluation of pain as an outcome, the differences between the various methods should be taken into account.

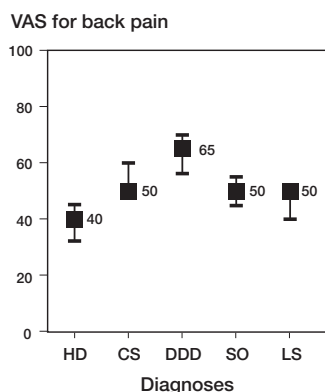


Figure 3. Preoperative pain graded on the VAS scale regarding back pain (means and 95% CI) related to diagnosis. Abbreviations: See page 3.

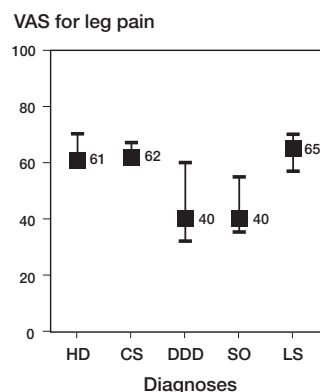


Figure 4. Preoperative pain graded on the VAS scale regarding leg pain (means and 95% CI) related to diagnosis.

Table 3. Patient satisfaction vs. other pain outcomes. The self graded patient satisfaction 12 months after surgery has been tested regarding correlation to other types of pain outcomes

	Patient satisfaction at follow-up
Perceived change (direct question)	
Back	0.60
Leg	0.60
VAS (absolute)	
Back	0.57
Leg	0.51
VAS (differential)	
Back	0.35
Leg	0.35
Use of pain killers	
(Absolute)	0.45
(Differential)	0.29

The correlation (Spearman's rho) was significant with $p < 0.01$ in all instances. Differential value = postoperative value minus preoperative value on the VAS scale. The patient satisfaction at 12 months postoperatively correlated best to the direct question on perceived change both regarding back and leg pain

Paper IV. SF-36 scores in degenerative lumbar spine disorders: analysis of prospective data from 451 patients

Study design. Retrospective analysis of prospectively collected cohort data.

Background. When using Health-Related Quality of Life (HRQoL) in assessing outcomes of treatment, normative data for different diagnoses

are needed to allow comparisons between existing and future studies.

Objective. Aim of this study is to determine the SF-36 scores in patients with lumbar spine problems scheduled for surgical treatment.

Methods. Between 1998 and 2002 477 patients undergoing elective lumbar surgery at the spine section of the Orthopedics department of the University of Lund were included in the protocol for the Swedish Lumbar Spine Registry which since 1998 contains the Swedish version of the SF-36 questionnaire. 451/477 patients (95%) returned the forms and their data will be presented.

Results. All baseline characteristics of patients included in the study are shown in Table 4.

Preoperative SF-36 scores were significantly lower than in the normal population and in patients with non specific Low-Back Pain (LBP) (Figure 5).

Significant but not very strong correlation was found between SF-36 scores and other preoperative clinical variables.

SF-36 profiles have been compared between several subgroups, categorized according to diagnosis or demographic characteristics, but no single strong correlation was noted in the 8 domains. Some SF-36 domains showed a possible discriminating pattern when analysed separately for the five diagnoses (Figure 7). Among other preoperative variables studied, age showed a similar discriminating performance, whereas duration of back and leg pain was significantly shorter only for HD (median = 9 months), compared to all other diagnoses (median = 24 months) ($p < 0.01$).

Table 4. Characteristics of study population

		Total (n = 477)	Responders (n = 451)	Nonresponders (n = 26)	Asymp. Sig. (2 tailed) ^a
Sex ^b	Male	240 (50.3)	224 (49.7)	16 (61.5)	0.240
	Female	237 (49.7)	227 (50.3)	10 (38.5)	
Age ^c		52 (50–55)	52 (50–54)	55 (47–71)	0.211
Duration (months) of	back pain ^c	17 (15–19)	18 (15–19)	15 (7–53)	0.664
	leg pain ^c	15 (14–18)	15 (13–18)	17 (7–40)	0.561
Smoking ^b	Yes	122 (25.6)	119 (26.4)	3 (11.5)	0.211
	No	347 (72.7)	329 (72.9)	18 (69.2)	
	Missing	8 (1.7)	3 (0.7)	5 (19.2)	
Diagnosis ^{b, d}	HD	150 (31.4)	144 (31.9)	6 (23.1)	0.531
	CS	130 (27.3)	121 (26.8)	9 (34.6)	
	DDD	63 (13.2)	57 (12.6)	6 (23.1)	
	SO	54 (11.3)	52 (11.5)	2 (7.7)	
	LS	57 (11.9)	55 (12.2)	2 (7.7)	
	Other	23 (4.8)	22 (4.9)	1 (3.8)	

^a Mann-Whitney; ^b Values are n (%); ^c Values are median (95% CI);

^d Abbreviations, see page 3 and Other: Miscellaneous and combined diagnoses.

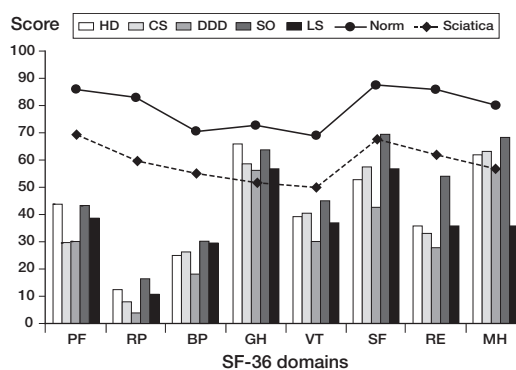


Figure 5. SF-36 profiles for the 5 diagnostic categories (bars) compared with normative values for the Swedish normal and sciatica population (lines). Abbreviations, see page 3.

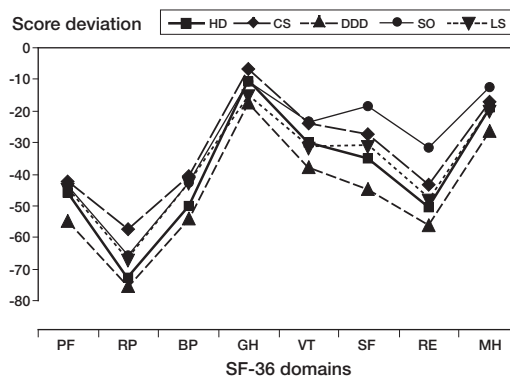


Figure 6. Mean deviations from the norms of the Swedish general population on the eight SF-36 scales for the five diagnostic subgroups, preoperatively.

Conclusion. HRQoL reported by patients scheduled for lumbar spine surgery showed a pronounced reduction compared to normal and LBP population. The difference was most pronounced concerning Physical Function. Role Physical and Bodily Pain had also very low values, whereas less difference appeared in mental domains (Figure 5). Normative SF-36 values provided may be used as a benchmark in future studies.

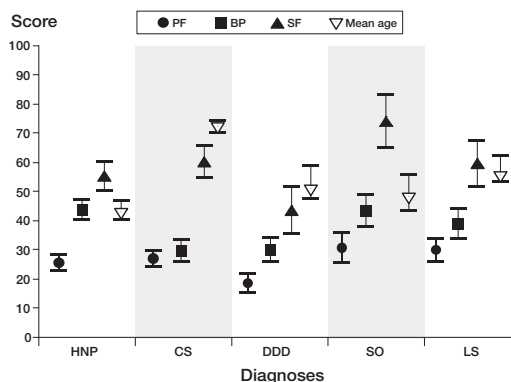


Figure 7. Mean values (with 95% CI) of the PF, BP, SF domain and median age for the 5 diagnostic categories.

Paper V. SF-36 for outcomes assessment of spine surgery

Study design. Retrospective analysis of prospectively collected cohort data.

Background. The most used generic instrument for HRQoL evaluation in spine pathologies is SF-36. There is no standard definition of important differences in spine surgery patients.

Objective. To describe the use of SF-36 in assessing 1-year outcomes of patients operated on for lumbar spine disorders.

Methods. Between 1998 and 2002, persons undergoing elective lumbar surgery at the spine section of the Orthopaedics Department of the University of Lund were included in the protocol. Data were collected preoperatively and after 1-year follow-up.

In addition to SF-36 profiles, norm-based scoring and improvement, responsiveness statistics and “k-means” cluster analysis were performed. Minimal clinically important differences (MCIDs) for the physical domains were described using mean differential values (postoperative values *minus* preoperative values) for SF-36 profiles and VAS scores. Comparison between preoperative and postoperative responses to the walking distance item enabled an *a posteriori* classification into 4 categories (worsened, unchanged, improved and no limitations, for those who reported over 1 km walking distance pre- and postoperatively). Self-rated retrospective pain change and satisfaction were used

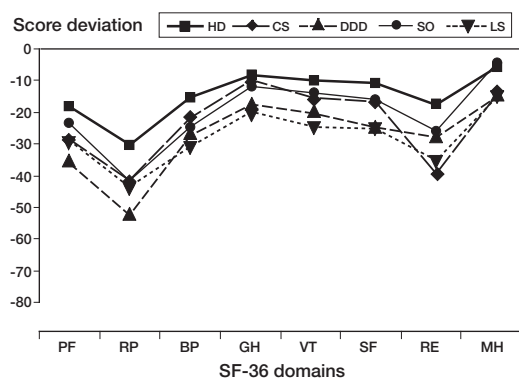


Figure 8. Mean deviations from the norms of the Swedish general population on the eight SF-36 scales for the five diagnostic subgroups, postoperatively. Abbreviations, see page 3.

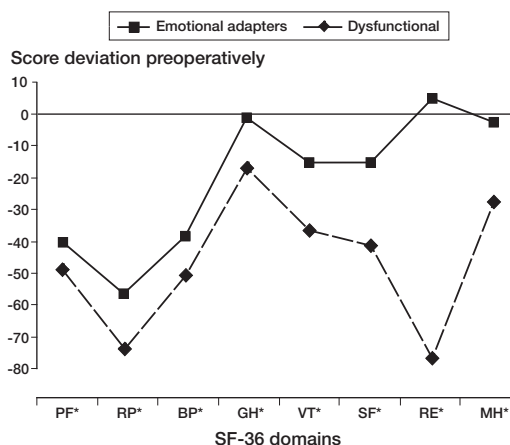


Figure 9. Mean deviations from the norms of the Swedish general population on the 8 SF-36 scales for the two identified clusters, preoperatively. * indicates a significant (Mann-Whitney, $p < 0.01$) difference between the two clusters.

to categorize data into clinically meaningful subgroups. Finally, responsiveness of relevant variables was investigated using 4 different methods as suggested by Walters and Brazier (2003)

Results. 351/397 patients (88%), median age 52 (13–86) were included: 29.1% had lumbar disc herniations (39% of which underwent microscopical discectomies, 35% open, and 21% APLD) 27.6% central and 13.7% lateral stenosis, 11.4% isthmic spondylolysis and 12.5% degenerative disc disease. At the 1-year follow-up all groups showed improvement in all SF-36 domains except General Health (GH). Taking in to account individual differences from the age- and sex-matched Swedish normative profiles, mean values, although improved, remained below the norm postoperatively (Figure 8).

Cluster analysis classified the patients into two groups. Group 1 (“emotional adapters”) had low measures on physical variables but were close to norm on the mental scales. Group 2 (“dysfunctional”) had scores much lower than the norm on all scales (Figure 9). Both groups showed a marked postoperative improvement in all physical domains except general health, the dysfunctional group also demonstrated an improvement in the mental domains, although they remained below the normal scores (Figure 10).

Conclusions. SF-36 scores from this database may serve as a source of reference for future studies on specific lumbar conditions. Data show

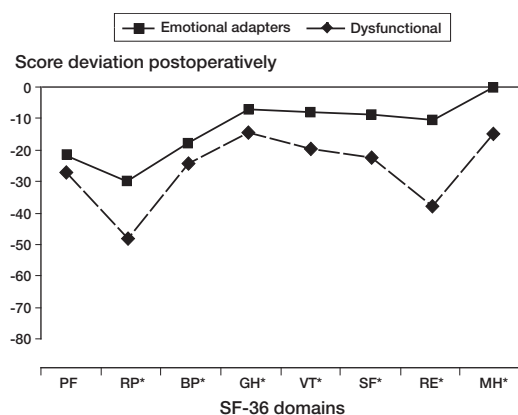


Figure 10. Mean deviations from the norms of the Swedish general population on the 8 SF-36 scales for the two identified clusters, postoperatively. * indicates a significant (Mann-Whitney, $p < 0.01$) difference between the two clusters.

improvement following surgery although other factors or natural history could contribute to this improvement.

Paper VI. SF-36 profiles before and one year after spinal stenosis surgery – a prospective comparison of two techniques in two nations

Study design. Retrospective comparison of two age- and sex-matched prospective cohort studies.

Background. Both a FDA study on X STOP®, a new implant for minimally invasive interspinous decompression, and the Swedish National Lumbar Spine Registry utilise SF-36 before surgery and after 12 months as an outcome measure.

Objectives. To investigate the utility of SF-36 in comparing two surgical techniques for spinal stenosis (X STOP® decompression, X-group, and traditional decompressive surgery, D-group), in two countries.

Methods. Patients aged >50 years with spinal claudication and scheduled for surgery were included. The X-group ($n=90$) was operated on in one of 9 US centers in a prospective FDA study. The D group was created to match age- and sex-distribution of the US sample. Out of 121 patients who had been operated on for lumbar central spinal stenosis in a single unit and had completed the 1-year follow-up in the Swedish national lumbar

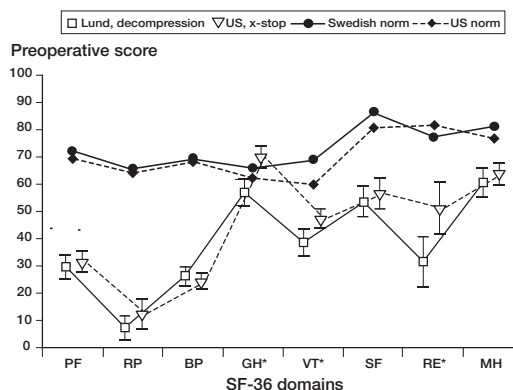


Figure 11. Preoperative SF-36 profiles (mean and 95% CI) for the D- and the X-group compared with normative values for the Swedish and US population. * indicates a significant (Mann-Whitney, $p < 0.01$) difference between D- and X-group.

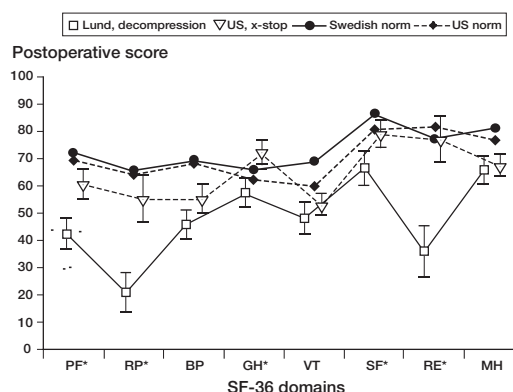


Figure 12. Postoperative SF-36 profiles (mean and 95% CI) for the D- and the X-group compared with normative values for the Swedish and US population. * indicates a significant (Mann-Whitney, $p < 0.01$) difference between D- and X-group.

spine register protocol, 90 patients were chosen from the database to form the matched control group.

Both patient groups completed the SF-36 prior to surgery and at 1-year follow-up.

Individual SF-36 scores were obtained for each questionnaire, and individual differential (postoperative *minus* preoperative) values from respective normative national age- and sex-matched scores were calculated on each occasion.

Results. The preoperative SF-36 scores for X- and D-groups showed similar mean values except

for General Health (GH), Vitality (VT) and Role Emotional (RE) domains, which were significantly higher for the X group. The postoperative values were statistically higher for the X group in the Physical Function (PF), Role Physical (RP), GH, Social Function (SF) and RE domains. A significant improvement was obtained for both groups in all domains except GH. Improvement in RE was significant only in the X-group. Higher values of improvement were reported in the X-group compared to the D group in the PF and RP domains

Conclusion. The use of standardised outcome measures allows international comparisons,

although caution should be used in the interpretation of differences. The patients selected for surgery for spinal stenosis in the US and Sweden seemed to have a comparable health related quality of life as judged by the preoperative SF-36 scores, although the preoperative differences in the General Health domain suggest the possibility of different prevalence of comorbidities in the two groups. Both surgical treatments improved HRQoL as described by SF-36 one year postoperatively. The present study design has no capability to rule out other confounding factors as an explanation of differences in favour of the X-group.

Reference tables

Frequencies of diagnoses and procedures in the whole sample 1993–2003 (paper III, IV and V)

Table 5. Distribution of patients in the five diagnostic groups in percentage

Diagnosis	Total	1993–1997	1998–2003
Lumbar disc herniation	42.6	50.6	33.4
Central stenosis	21.6	16.9	26.9
Degenerative/ segmental pathology	6.8	3.4	10.7
Spondylolisthesis	9.4	8.1	11.0
Lateral stenosis	14.1	14.0	14.3

Table 6. Elective surgical procedures performed in percentage

Diagnosis	Total	1993–1997	1998–2003
Conventional discectomy	17.0	20.0	13.7
Microscopic discectomy	12.7	13.0	12.3
Percutaneous discectomy	11.4	15.9	6.2
Decompression	35.6	30.3	41.7
Instrumented anterior fusion	1.8	1.4	2.2
Uninstrumented posterior fusion	5.0	4.7	5.3
Instrumented posterior fusion	5.7	5.4	6.2
Decompression + uninstrumented posterior fusion	0.6	0.8	0.5
instrumented posterior fusion	8.4	7.0	10.1

SF-36 reference values (paper IV and V)

Table 7. Mean SF-36 scores for the whole sample

WHOLE SAMPLE	n (%)	Age	Males (%)	PF	RP	BP	GH	VT	SF	RE	MH
Preoperative scores	451 (100)	52 (50–55)	49.7								
Mean score				37.27	10.44	25.81	60.93	38.74	55.40	35.88	62.25
SD				20.64	24.48	15.89	24.04	22.46	27.62	43.59	22.58
95% CL, lower				35.36	8.18	24.34	58.70	36.67	52.85	31.85	60.15
95% CL, upper				39.18	12.71	27.28	63.15	40.82	57.96	39.92	64.34
Postoperative scores	351 (100)	52 (50–55)	49.6								
Mean score				57.83	36.23	50.14	60.40	52.32	69.59	54.89	71.05
SD				27.22	42.34	27.73	25.33	26.74	30.09	45.85	24.03
95% CL, lower				54.97	31.77	47.22	57.74	49.50	66.43	50.08	68.52
95% CL, upper				60.69	40.69	53.06	63.06	55.13	72.75	59.70	73.58
Preoperative differences from norm	451 (100)	52 (50–55)	49.7								
Mean score				-45.76	-67.41	-46.20	-11.30	-28.86	-31.89	-47.21	-18.71
SD				20.33	26.46	16.53	23.58	22.46	27.91	43.25	22.85
95% CL, lower				-47.65	-69.86	-47.73	-13.48	-30.94	-34.47	-51.22	-20.82
95% CL, upper				-43.88	-64.96	-44.67	-9.11	-26.78	-29.31	-43.21	-16.60
Postoperative differences from norm	351 (100)	52 (50–55)	49.6								
Mean score				-25.19	-41.88	-22.06	-11.74	-15.61	-17.73	-28.27	-9.91
SD				25.97	41.99	27.79	25.41	27.03	30.30	44.34	24.63
95% CL, lower				-27.91	-46.29	-24.97	-14.41	-18.44	-20.91	-32.92	-12.50
95% CL, upper				-22.46	-37.47	-19.14	-9.08	-12.77	-14.55	-23.61	-7.33

Age is median and 95% CI, Mean score is the group mean, SD the standard deviation. 95% CL is the 95% confidence limits.

Table 8. Mean SF-36 scores for persons who underwent elective lumbar spine surgery for lumbar disc herniation

HD	n (%)	Age	Males (%)	PF	RP	BP	GH	VT	SF	RE	MH
Preoperative scores	144 (32)	41 (39–45)	55								
Mean score				43.77	12.67	25.17	65.87	39.14	53.21	36.11	62.06
SD				20.36	27.26	16.45	23.45	21.52	26.84	43.45	21.88
95% CL, lower				40.42	8.18	22.46	62	35.6	48.79	28.95	58.46
95% CL, upper				47.12	17.16	27.88	69.73	42.69	57.63	43.27	65.67
Postoperative scores	102 (29)	42 (39–45)	52								
Mean score				71.13	50.25	58.08	66.44	57.79	75.25	66.67	74.23
SD				23.22	43.59	28.72	25.42	26.71	27.92	42.73	23.26
95% CL, lower				66.57	41.68	52.44	61.45	52.55	69.76	58.27	69.66
95% CL, upper				75.69	58.81	63.72	71.44	63.04	80.73	75.06	78.79
Preoperative differences from norm	144 (32)	41 (39–45)	55								
Mean score				-46.24	-73.25	-50.28	-10.70	-30.27	-35.50	-51.03	-19.04
SD				19.80	27.39	16.68	23.43	21.20	26.93	43.49	21.86
95% CL, lower				-49.50	-77.76	-53.02	-14.56	-33.76	-39.94	-58.20	-22.64
95% CL, upper				-42.98	-68.74	-47.53	-6.84	-26.78	-31.07	-43.87	-15.44
Postoperative differences from norm	102 (100)	42 (39–45)	52								
Mean score				-18.62	-35.41	-17.06	-9.94	-11.51	-13.38	-20.37	-6.87
SD				23.17	43.75	28.27	25.07	26.24	27.71	42.45	22.96
95% CL, lower				-23.17	-44.01	-22.61	-14.87	-16.66	-18.82	-28.71	-11.38
95% CL, upper				-14.07	-26.82	-11.51	-5.02	-6.36	-7.93	-12.03	-2.36

Age is median and 95% CI, Mean score is the group mean, SD the standard deviation. 95% CL is the 95% confidence limits.

Table 9. Mean SF-36 scores for persons who underwent elective lumbar spine surgery for central stenosis

CS	n (%)	Age	Males (%)	PF	RP	BP	GH	VT	SF	RE	MH
Preoperative scores	121 (27)	70 (67–71)	42								
Mean score				29.68	8.13	26.56	58.71	40.50	57.54	32.92	63.43
SD				20.37	21.04	15.58	23.06	22.50	27.12	44.54	23.50
95% CL, lower				26.01	4.34	23.76	54.56	36.45	52.66	24.9	59.19
95% CL, upper				33.35	11.91	29.37	62.86	44.55	62.42	40.94	67.68
Postoperative scores	97 (28)	69 (67–71)	41								
Mean score				43.03	21.01	45.39	57.82	48.30	67.27	37.46	66.57
SD				25.98	34.35	25.81	25.03	27.73	30.21	44.94	24.94
95% CL, lower				37.79	14.05	40.19	52.77	42.71	61.18	28.40	61.55
95% CL, upper				48.27	27.97	50.59	62.86	53.89	73.36	46.52	71.60
Preoperative differences from norm	121 (27)	70 (67–71)	42								
Mean score				-42.42	-57.22	-40.79	-6.97	-23.90	-27.26	-43.27	-16.48
SD				21.72	24.84	15.77	23.03	22.87	27.81	43.30	23.70
95% CL, lower				-46.33	-61.69	-43.63	-11.11	-28.01	-32.26	-51.07	-20.74
95% CL, upper				-38.51	-52.75	-37.96	-2.82	-19.78	-22.25	-35.48	-12.21
Postoperative differences from norm	97 (28)	69 (67–71)	41								
Mean score				-29.42	-44.96	-22.01	-8.02	-16.19	-17.65	-38.99	-12.80
SD				27.63	37.99	26.59	26.74	29.16	31.10	44.27	25.69
95% CL, lower				-34.99	-52.62	-27.37	-13.40	-22.07	-23.92	-47.92	-17.98
95% CL, upper				-23.85	-37.30	-16.66	-2.63	-10.31	-11.39	-30.07	-7.63

Age is median and 95% CI, Mean score is the group mean, SD the standard deviation. 95% CL is the 95% confidence limits.

Table 10. Mean SF-36 scores for persons who underwent elective lumbar spine surgery for degenerative disc disease

DDD	n (%)	Age	Males (%)	PF	RP	BP	GH	VT	SF	RE	MH
Preoperative scores	57 (13)	49 (45–56)	51								
Mean score				30.02	4.32	18.39	56.29	30.44	42.98	28.07	56.73
SD				15.87	13.67	11.98	26.18	20.66	26.73	41.69	19.87
95% CL, lower				25.81	0.65	15.21	49.34	24.96	35.89	17.01	51.36
95% CL, upper				34.23	7.98	21.56	63.24	35.92	50.07	39.13	62.10
Postoperative scores	44 (13)	47 (43–53)	55								
Mean score				53.57	31.40	49.70	56.66	53.37	62.78	59.09	70.14
SD				25.90	40.86	31.75	28.78	24.85	33.62	43.65	25.75
95% CL, lower				45.69	18.82	39.93	47.91	45.72	52.56	45.82	62.21
95% CL, upper				61.44	43.97	59.47	65.41	61.02	73.00	72.36	78.06
Preoperative differences from norm	57 (13)	49 (45–56)	51								
Mean score				-54.67	-75.32	-53.97	-16.94	-37.66	-44.64	-56.15	-26.27
SD				16.47	16.43	12.81	24.97	20.69	26.80	41.23	22.34
95% CL, lower				-59.04	-79.68	-57.37	-23.56	-43.15	-51.75	-67.09	-32.20
95% CL, upper				-50.30	-70.96	-50.57	-10.31	-32.18	-37.53	-45.22	-20.35
Postoperative differences from norm	44 (13)	47 (43–53)	55								
Mean score				-33.51	-51.61	-24.73	-17.72	-16.72	-25.47	-26.84	-12.76
SD				25.29	39.18	32.30	28.20	26.41	33.75	42.69	27.81
95% CL, lower				-41.20	-63.52	-34.56	-26.29	-24.75	-35.73	-39.82	-21.22
95% CL, upper				-25.82	-39.70	-14.91	-9.14	-8.69	-15.21	-13.86	-4.31

Age is median and 95% CI, Mean score is the group mean, SD the standard deviation. 95% CL is the 95% confidence limits.

Table 11. Mean SF-36 scores for persons who underwent elective lumbar spine surgery for spondylolisthesis

SO	n (%)	Age	Males (%)	PF	RP	BP	GH	VT	SF	RE	MH
Preoperative scores	52 (12)	46 (41–53)	35								
Mean score				43.39	16.35	30.13	64.00	45.00	69.71	53.85	68.15
SD				19.56	32.04	18.03	25.28	24.71	28.48	44.85	20.99
95% CL, lower				37.95	7.43	25.12	56.96	38.12	61.78	41.36	62.31
95% CL, upper				48.84	25.27	35.15	71.03	51.88	77.64	66.33	74.00
Postoperative scores	40 (11)	46 (41–53)	33								
Mean score				64.66	43.96	51.20	65.34	54.92	75.00	66.67	78.43
SD				22.71	46.61	26.62	20.86	24.36	25.63	44.66	15.83
95% CL, lower				57.40	29.05	42.69	58.67	47.13	66.80	52.38	73.36
95% CL, upper				71.93	58.86	59.71	72.02	62.71	83.20	80.95	83.49
Preoperative differences from norm	52 (12)	46 (41–53)	35								
Mean score				-43.57	-65.84	-42.83	-10.55	-23.45	-18.51	-31.83	-12.47
SD				20.19	31.67	16.65	23.83	24.53	28.24	44.32	20.79
95% CL, lower				-49.20	-74.66	-47.47	-17.18	-30.27	-26.37	-44.16	-18.26
95% CL, upper				-37.95	-57.03	-38.20	-3.91	-16.62	-10.64	-19.49	-6.68
Postoperative differences from norm	40 (11)	46 (41–53)	33								
Mean score				-22.06	-38.03	-21.50	-9.06	-13.44	-13.19	-18.94	-2.21
SD				23.53	46.99	26.84	19.40	24.24	25.52	44.91	15.96
95% CL, lower				-29.59	-53.06	-30.08	-15.26	-21.20	-21.35	-33.31	-7.32
95% CL, upper				-14.54	-23.00	-12.91	-2.85	-5.69	-5.02	-4.58	2.89

Age is median and 95% CI, Mean score is the group mean, SD the standard deviation. 95% CL is the 95% confidence limits.

Table 12. Mean SF-36 scores for persons who underwent elective lumbar spine surgery for lateral stenosis

LS	n (%)	Age	Males (%)	PF	RP	BP	GH	VT	SF	RE	MH
Preoperative scores	55 (12)	53 (51–60)	64								
Mean score				38.82	10.91	29.45	56.84	37.12	57.05	35.76	61.25
SD				19.08	22.96	14.19	20.59	20.03	25.66	41.99	24.00
95% CL, lower				33.66	4.7	25.62	51.28	31.71	50.11	24.4	54.77
95% CL, upper				43.98	17.11	33.29	62.41	42.54	63.98	47.11	67.74
Postoperative scores	48 (14)	53 (51–60)	67								
Mean score				55.14	33.51	41.35	52.54	44.62	63.28	52.08	67.65
SD				25.83	40.70	24.93	23.42	27.38	30.45	47.59	27.48
95% CL, lower				47.64	21.69	34.12	45.74	36.67	54.44	38.26	59.67
95% CL, upper				62.65	45.32	48.59	59.34	52.57	72.12	65.90	75.63
Preoperative differences from norm	55 (12)	53 (51–60)	64								
Mean score				-44.58	-67.19	-42.70	-15.05	-31.22	-30.61	-48.12	-19.67
SD				20.12	24.55	15.63	21.39	19.87	26.05	43.30	23.65
95% CL, lower				-50.02	-73.82	-46.63	-20.83	-36.59	-37.66	-59.83	-26.06
95% CL, upper				-39.14	-60.55	-38.17	-9.27	-25.85	-23.57	-36.42	-13.28
Postoperative differences from norm	48 (14)	53 (51–60)	67								
Mean score				-28.06	-44.24	-30.58	-19.24	-23.72	-24.30	-31.59	-13.21
SD				24.97	41.12	24.71	22.79	26.91	30.85	46.25	27.33
95% CL, lower				-35.31	-56.18	-37.75	-25.86	-31.53	-33.26	-45.02	-21.15
95% CL, upper				-20.81	-32.30	-23.40	-12.62	-15.91	-15.34	-18.16	-5.28

Age is median and 95% CI, Mean score is the group mean, SD the standard deviation. 95% CL is the 95% confidence limits.

Responsiveness (paper V)

Table 13. Responsiveness statistics, their formulas, and relevant differential values and measures of variability used in the calculation of pain and functional outcomes

	SF-36 domains				
	VAS back	VAS leg	PF	RP	BP
A. Mean of the individual differential values ^a	23.98	29.16	21.09	26.02	24.10
B. Standard deviation of the differential values	36.05	35.52	26.36	44.02	29.34
C. Standard deviation of the preoperative values	27.97	27.61	20.64	24.48	15.89
D. Standard error of the differential values	1.32	1.39	1.41	2.35	1.57
E. Standard deviation of the postoperative values	30.26	31.47	27.36	42.34	27.73
F. Standard deviation of the differential values in patients who considered themselves stable	25.80	28.86	20.13	42.98	10.27
G. Standardised response mean (A/B)	0.67	0.82	0.80	0.59	0.82
H. Standardised effect size (A/C)	0.86	1.06	1.02	1.06	1.52
I. Paired T statistics (A/D)	18.21	20.91	14.99	11.07	15.39
J. Responsiveness statistics (A/F)	0.93	1.01	1.05	0.61	2.35

^a between postoperative and preoperative values

Reliability (paper II)

Table 14. Intra-class Correlation Coefficients for the outcome variables in the two samples

Variable	Preoperative sample			Postoperative sample		
	n	ICC ^a	95% CI	n	ICC ^a	95% CI
Preoperative duration (months)						
Back pain	60	0.85	0.77–0.91			
Leg pain	60	0.84	0.75–0.90			
Preoperative VAS for pain intensity						
Back pain	60	0.78	0.66–0.87	55	0.89	0.82–0.93
Leg pain	60	0.88	0.81–0.93	55	0.85	0.75–0.91
Preoperative SF-36 domain						
PF	60	0.82	0.72–0.89	59	0.86	0.78–0.92
RP	60	0.85	0.76–0.91	59	0.71	0.56–0.82
BP	60	0.69	0.51–0.80	59	0.86	0.78–0.92
GH	60	0.86	0.77–0.92	59	0.80	0.68–0.88
VT	60	0.67	0.46–0.80	59	0.87	0.79–0.92
SF	60	0.77	0.65–0.86	59	0.65	0.48–0.78
RE	60	0.67	0.51–0.79	59	0.62	0.43–0.76
MH	57	0.89	0.83–0.94	59	0.83	0.73–0.89

^a Two-way random model Intraclass Correlation Coefficient

Table 15. Agreement (weighted kappa) for the outcome variables in the two samples

Variables at 12 months	Preoperative sample			Postoperative sample		
	n	K ^a	SE ^b	n	K ^a	SE ^b
Back pain improvement ^c				54	0.74	0.06
Leg pain improvement ^c				53	0.85	0.04
Work situation	57	0.91	0.04	55	0.91	0.04
Use of pain-killers	59	0.85	0.06	53	0.87	0.05
Walking distance	58	0.81	0.05	55	0.77	0.07
Patient's satisfaction				51	0.83	0.09
SF-36 domain						
RP	60	0.76	0.08	59	0.64	0.08
RE	60	0.64	0.10	59	0.53	0.09

^a Weighted kappa; ^b Standard error; ^c self-rated

MCIDs (paper V)

Table 16. Differential values for several pain outcomes categorised by patients' rating of postoperative pain improvement. Values are mean (95% CI)

Postoperative minus preoperative rating	Patient self-reported separate outcome in the protocol			
	12-month back pain improvement		12-month leg pain improvement	
	Somewhat better	Much better	Somewhat better	Much better
Change in VAS for intensity of back pain	8 (1 – 14)	34 (29–40)	–	–
Change in VAS for intensity of leg pain	–	–	19 (12–27)	39 (33–45)
Change in SF-36 Bodily Pain subscale	10 (7–14)	34 (30–37)	16 (12–22)	29 (24–34)

Table 17. Differential values for several pain and functional outcomes categorised by patients' rating of postoperative improvement. Values are mean (95% CI)

Postoperative minus preoperative rating	Patient self-reported separate outcome in the protocol				
	Much better	Somewhat better	Unchanged	Somewhat worse	Much worse
Change in VAS for intensity of back pain	42 (38 – 47)	14 (6 – 21)	2 (-5 – 8)	-5 (-8 – 17)	-9 (-32 – 14)
Change in VAS for intensity of leg pain	46 (42 – 51)	19 (11 – 26)	9 (2 – 17)	3 (-5 – 12)	-7 (-26 – 12)
Change in SF-36 subscale					
Physical Function	36 (32 – 40)	18 (13 – 23)	9 (3 – 13)	3 (-5 – 10)	2 (-8 – 12)
Role Physical	48 (41 – 55)	14 (6 – 22)	16 (6 – 25)	0 (-11 – 11)	-10 (-26 – 6)
Bodily Pain	45 (40 – 49)	16 (12 – 21)	10 (5 – 13)	-2 (-11 – 6)	-4 (-12 – 3)

Table 18. Crosstabulation between pain outcomes categorised by patients' rating of postoperative improvement and mean score improvement across the five main diagnostic subgroups. Values are median

Diagnosis	Postoperative minus preoperative rating	Patient self-reported separate outcome in the protocol			
		Back pain		Leg pain	
		Somewhat better	Much better	Somewhat better	Much better
HD	Change in VAS for intensity of back pain	4	23	–	–
	Change in VAS for intensity of leg pain	–	–	9	58
	Change in SF-36 Bodily Pain subscale	10	41	19	40
CS	Change in VAS for intensity of back pain	8	38	–	–
	Change in VAS for intensity of leg pain	–	–	32	47
	Change in SF-36 Bodily Pain subscale	9	31	10	19
DDD	Change in VAS for intensity of back pain	21	49	–	–
	Change in VAS for intensity of leg pain	–	–	19	24
	Change in SF-36 Bodily Pain subscale	10	34	15	18
SO	Change in VAS for intensity of back pain	15	43	–	–
	Change in VAS for intensity of leg pain	–	–	7	43
	Change in SF-36 Bodily Pain subscale	10	41	19	40
LS	Change in VAS for intensity of back pain	14	38	–	–
	Change in VAS for intensity of leg pain	–	–	15	27
	Change in SF-36 Bodily Pain subscale	10	31	10	32

Discussion

This dissertation focuses on outcome assessment after lumbar spine surgery, analyzed mainly by means of utilizing prospectively collected register data. When leaving a known track for a new one, it must be taken into account that it may take longer to reach the destination. Some study designs are more straightforward for journals (and peer reviewers) and some manuscripts find acceptance for publication easier. Previous theses on outcome from the Scandinavian literature share this difficulty, with less than 50% of manuscripts published before dissertation on average (Roos 1999, Robertsson 2000, Dunbar 2001, Hägg 2002). Methodological studies certainly meet with more difficulty in passing the peer-review in clinical journals and in the discussion it is constantly requested to interpret results within the boundaries of the design, strengths and weaknesses of the study.

So this discussion will start by addressing first of all the particular type of study designs and possible weaknesses of the single papers, and then proceed to a more general content analysis.

The project of the systematic review started in 1999 and was proposed at different institutions and individuals for review: the initial reactions were rather diffident as it did not fit in any well-established research protocol. Even though probably considered “crazy” or impossible to accomplish, the idea finally got accepted as an “exotic” touch and published together with other more “serious” articles. Looking back at this study after many years and having had the chance to perform other more traditional systematic (Cochrane) reviews in the meantime, the strongest limitations of paper I is the impossibility to be sure to have been systematic enough. Outcome instruments are rarely identifiable in Medline, even when they constitute the main object of a paper: the proposal to mark records with an “outcome” tag as for RCTs or meta-analyses has produced no consequences, even though other websites have emerged that gather references on different outcome instruments or give updated information (e.g. existing versions and languages) on single ones (Fairbank 2004). Despite the

impressive number of instruments retrieved, the publication in one of the most influential journals in the field, and the substantial approval that it has received by many experts, this paper has failed in preventing new outcome instruments to be proposed for LBP patients. This is probably a task that should be taken over by the international spine societies, which can require uniformity and standardisation in the papers they accept for presentation or publication.

The test-retest protocol used in paper II is rather standardized, and does not present particular obscurity in the interpretation of its methodology. The only uncontrolled event was that not all patients complied with the 1 week recommendation and answered the questionnaire after different time spans. Rather than simply excluding these cases from the analysis, the *recycling* principle mentioned in the introduction was applied, trying to maximize the usefulness of collected information, and subgroups analyses were performed for different time spans between test and retest.

Paper III, IV and V represent the real field of application of the *ecological* approach of *recycling* (otherwise lost) complex data. The (few) strengths and (many) weaknesses of this study design are already obvious in its formal definition: a retrospective analysis of prospectively collected data with some lost to follow-up. Hypothesis testing is the base of experimental studies. This was rather an observational study, so no inference on the expected results was performed. A retrospective analysis, implies that no control can be made on the quality and quantity of data collected. In this case the methodological weakness was compensated by the fact that a formally explicit prospective registration protocol existed, although it varied over time. However, two general limitations could not be avoided: the first one was intrinsic in the population analysed, namely the inclusion criterion being that patients already had been selected for operation. Thus, findings from these studies are not readily extensible to the population of common low-back sufferers. The second general limitation

is the absence of a comorbidity measure, which could help interpret the results of a generic HRQoL instrument as SF-36. Several techniques, such as the use of norm-based scoring, were applied to overcome this potential limitation: controlling for age and sex, some of the co-morbidity should be already taken into account. A threat to the validity of the results is located in the loss of cases: apart from those who never entered the protocol because of change of routines at the ward and those who were lost-to-follow-up for various reasons, the most frequent event was that of incomplete data entering among returned questionnaires. In an observational design, being close to 100% of the expected population is crucial to ensure that conclusion drawn from analysis of collected data are not biased. Although there was no reason to believe that any selection bias had significantly acted on the sample of responders, a preliminary check was always performed to ensure that the samples did not show significant differences with the whole as regards demographics and preoperative variables. However, if these had been outcome studies instead of studies on outcome the follow-up rate would have probably been insufficient. Conversely, since the point was not to demonstrate the effectiveness of surgical techniques, but to describe methodological properties of different ways of measuring outcomes after lumbar spine surgery, missing data should not hamper the validity of the results presented. Rather than trying to act on the data to increase their comprehensiveness, using statistical techniques or protocol changes to put reality *under control*, the ecological approach suggested to *cope with the complexity* and incompleteness of the data without wasting it. The inability to fully describe the population was compensated by a strategy which tried to be closer to the reality of everyday clinical outcome assessment..

In paper VI, the methodology used in comparing two surgical techniques in two nations might seem inappropriate. In fact, the direct comparison of two treatment strategies in the same type of patients can only be addressed conclusively by a RCT. Once again, this was not the point of the study, which rather tried to explore the possibility to test the usefulness of international standardisation in outcome collection from a methodological point of view. Retrospectively comparing existing prospec-

tive data can give a preliminary idea on how the two samples behave, even though any difference cannot be attributed to the technique because it is impossible rule out other potential confounding factors. This is the final task of a well-designed and conducted RCT, but the first impression achieved by using the *recycling* approach is undoubtedly faster and cheaper, and can help in properly planning a subsequent RCT.

As for the results, the Swedish Lumbar Spine Register protocol is designed to achieve discrimination between groups, possibly large groups of patients. Data presented in paper II and comparison with similar data in the literature (Dunbar et al. 2001, Roos et al. 1998, Ostelo et al. 2004, Hagberg et al. 2004) corroborate the conclusion that all the outcome variables under study seem reliable enough, although some might be more so than others, to detect the rather pronounced improvement which is usually seen postoperatively in this type of patients (Zanoli et al. 2004). On more speculative grounds, one could say that to allow for a good generalizability of the results, postoperative questionnaire completion 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. In general, postoperative results seemed to be more stable and less affected by a longer time span between the two measurement. This might be due to the fact that the achieved improvement which we have shown elsewhere (Paper V) is not prone to day to day modifications as 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. A possible explanation might be that in 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) might produce less reliable information.

It is typical of the western medicine tradition to try to split the continuum of real life conditions into quanta – the diagnoses. Diagnoses are in a way abstract categorisations that we need to communicate within the profession and to standardise clinical practice, in order to be able to use the available scientific evidence. Very often, especially at multi-

disciplinary meetings, physicians with different specialty training have difficulties 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 might believe that there is no need to use the term fibromyalgia for patients who simply have (psychological) pain everywhere, physical therapists and rehabilitation specialists could argue that most adults have degenerative disc disease without needing a surgical intervention, and anyone who has not had a certain subspecialty training might have difficulties in accepting the centralization phenomenon or the existence of teeth-related low-back pain.

There is probably a little bit of truth in all 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 use of norm-based scoring (Ware 2000; Ware et al. 2000) allows for an immediate visualisation of the concept (Figure 6) 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 which limits its utility. One exception, however, comes from the musculoskeletal field (Strömbeck et al. 2000). Authors presented this type of information in a group of patients with rheumatologic diagnoses: Sjögren's syndrome and rheumatoid arthritis had significantly lower values than the norm, but only fibromyalgia showed a very pronounced gap in the first three physical domains which is similar to our sample. Interestingly, these rheumatologic conditions seemed to affect also GH scores and the mental domains in a rather homogeneous way, whereas in our sample the effect on physical domains was more instantly recognizable. 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. Graphically (Figure 5) the differences from normative values as reported in the Swedish SF-36 manual (Sullivan et al 1994) are substantial, especially in the first three physical domains. In fact they are lower than previously reported even

outside the normative studies for LBP (McKinnon et al. 1997) and for other musculoskeletal conditions (Gartsman et al. 1998; Stucki et al 1995). 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. Similarly, diagnostic workup 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 between workers and non-workers (sick leave, unemployed, retired). It is interesting to underline 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 National Spine Network on neck pain patients (Hee et al. 2002).

Even the findings from cluster analysis suggest that surgically selected patients are different. The two distinct clusters, defined by the profiles across the eight subscale scores of the SF-36 in Figure 9, groups seemed comparable to two of the three groups described by Fanciullo et al (2003) in accordance to previous reports (Jamison et al. 1987, Turk & Rudy 1988). Even the terms used in their study, "emotional adapters" and "dysfunctional", seemed quite appropriate in describing the two clusters identified by our analysis. The fact that a "highly functional" group as described in the above mentioned study was absent in our sample is another proof of the difference in the population of the two studies, the "highly functional" most likely not being appropriate candidates for surgery. For the same reasons demographics of the clusters in our sample are not completely super imposable to those of Fanciullo et al. Our data also confirm that the percentage of patients in each cluster varies with diagnosis. (Paper V). A possible explanation for this finding is the selection made before establishing an indication for surgery (Table 1), which probably succeeded in choosing subgroups

of patients whose perceived HRQoL is somewhat homogenous and so strongly influenced by the disease under treatment, that other variables become irrelevant. It can therefore be considered as a further confirmation of the diagnostic process. The two clusters remained significantly different at the one-year follow-up (Figure 10), although differences were reduced by a greater improvement in the mental domain by the dysfunctional groups. Clusters did not predict other outcome differences after 12 months.

Preoperative VAS values showed a clear cut pattern between various clinical diagnoses. Clinically consistent is also 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 disc disease patients. The most clear-cut distinction occurs in the BP domain, where HD patients seem to experience significantly more pain than the others. How can, for instance, young HD patients "know" (in mean) that their back pain is less than that of older CS patients? Why do old CS patient rate their back pain to be less pronounced than younger DDD patients do? Most patients usually have no or limited previous experience of other types of back pain and still they seem to rate their back pain and HRQoL in a consistent manner within diagnostic groups. We are not proposing the surveys as diagnostic tool, nor do we pretend that our limited findings can be generalised. In this respect, one could also note that age of the patients at presentation also shows a similar discriminative ability. We rather interpret this finding as an implicit confirmation of the selection process of surgical patients in our clinic: for example one would expect HD patients to be younger than CS ones, and this is confirmed by our observation. As a matter of fact, 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 such conclusions. However, from a personal and practical point of view, this clinical consistency was most rewarding information yielded by the data analysis of our sample.

Despite the statistically significant improvement in all physical domains and in some mental

ones across the five diagnostic groups (Table 7), norm-based scoring in figure 6 shows that individuals included in this study still do not reach normal profiles for age- and sex-matched population after one year (Figure 8). However, if we look at subsequent tables in the search for clinically meaningful values, we can say that for certain diagnoses (i.e lumbar disc herniation) HRQoL is within the range of clinically irrelevant differences from normal for mental scores and at borderline level for a moderate difference for physical scores, whereas for DDD group, which has probably the worst profile, HRQoL shows moderate to marked differences from the normal population throughout all domains. As stated before, this is a very special subgroup of profoundly disabled patients, which could compare well with the chronic low-back pain patients selected for randomization between surgery and conservative treatment in the recent surgical RCTs on chronic LBP (Brox et al 2003, Fairbank et al 2004, Fritzell et al 2001). It is important to underline that this is not a selected sample, but rather represents "a real life scenario", and is representative for patients who are treated in a specialised Spine Unit in a University Hospital. Possibly, cases operated here are more severe than those operated in a non-teaching Hospital (Rosenthal et al. 1997). Certainly, the long pre-operative duration of symptoms observed rather represents the effect of long waiting lists, than the application of time-related selection criteria or non-surgical treatment methods. However, caution should be applied when extrapolating this data to systems with shorter waiting lists: time might have in a way selected a cohort of patients with inferior natural history, but long-lasting untreated symptoms might also have increased their expectations toward a treatment which has a dramatic placebo effect. Anyway, our data is possibly more reflective of health care in general than samples selected for clinical trials such as presented in the literature. And even patients in the DDD group, although severely affected by the disease in their HRQoL from the start, showed the most pronounced effect. It is obvious that – in the absence of a control group – it is impossible to determine how much of the observed improvement could have occurred without surgery (due to natural history, concomitant therapies or placebo effect). The initial score

can influence the magnitude of change measured in a cohort of patients (lower scores allowing more room for improvement), but effect sizes can also reflect the quality of the intervention under study. The global effect of spine surgery in our sample is similar to effect sizes of very successful orthopedic interventions such as hip arthroplasty (Ostendorf et al. 2004), knee arthroplasty (Dunbar 2001; Dunbar et al 2004), knee arthroscopy (Roos 1999) and surgery for hip fractures (Tidemark et al. 2004), when analysed in severely impaired patients. Although the degree of recovery directly attributable to surgical treatment cannot be determined with the present study design, observational data presented in this study show an improvement in SF-36 and VAS scores after surgical treatment of degenerative conditions of the lumbar spine in carefully selected patients. It is important to integrate the information from the present paper to explain to persons severely affected by spinal conditions that it might not be always realistic to expect that surgery will yield a "normal" quality of life. This might be explained by the long preoperative duration of symptoms, having produced a chronic dysfunction in the physical aspects only for some patients and in both physical and mental components of HRQoL in others, as illustrated by cluster analysis. But it might also reflect the fact that present surgical technology in lumbar spine does allow for a complete functional recovery, even though patients in general show a significant improvement in most outcome measures used. Finally, co-morbidities could also contribute to the difference, even though this to some extent is contradicted by the fact that the GH score was the most stable and close to the norm among all the "physical" domains. Improved selection of surgical candidates in the future, possibly with the help of properly designed RCTs, might be the key to appropriate clinical decisions in order to maximize the effect for those who can benefit from spinal surgery and minimize the risks and the costs of a failed back syndrome for the others.

Paper VI focused on lumbar spinal stenosis, which is a well defined entity today regarding clinical presentation and radiographic appearance. In its typical form, the patients experience progressive leg symptoms (pain, numbness, tingling, weakness) on walking and the symptoms are relieved

by flexion of the spine. MRI or CT demonstrate a pronounced reduction of the dural sac area on axial images.

The natural course is not seldom benign (Johnsson et al. 1992), however in patients with long-standing problems, refractory to conservative treatment, good outcomes of decompressive procedures have been presented (Amundsen et al. 2000, Atlas et al 1996, Atlas et al 2000, Johnsson et al. 1991). Still, decompressive surgery/laminectomy yields a number of surgical and postoperative complication such as dural laceration, nerve root injury, cauda equina syndrome and superficial and deep wound infection, and the complication rate increases when concomitant segmental fusion is added (Strömqvist et al 2001; Gibson et al., 2000). Mortality shortly after surgery seems to be doubled when fusion is included in the decompressive procedure (Jansson et al 2003) and complications in open surgery occur to some extent (Deyo et al. 1992, Herno et al. 1993, Strömqvist et al 2001). Thus, there would seem to be a place for technical refinements and less invasive surgery and new methods have emerged on the market, like the device presented in paper VI. A study comparing X STOP surgery to epidural steroid injection in lumbar one or two level stenosis is completed with a 2 year follow up, and preliminary reports have been promising (Zucherman et al. 2004). The next step would logically be to compare the X STOP procedure with standard surgical decompression, and for this a prospective randomized study has been initiated in 2002. However it will take time to complete enrolment and the protocol requires a two-year follow-up of all patients included before definite conclusion can be drawn. Meanwhile, data presented in Paper VI should be regarded as a *first impression*, obtained with a concomitant cohort acting as a control group. Limitations of this study design have already been highlighted, and results should be looked at mainly from a methodological point of view. Patients selected for surgery for spinal stenosis in the US and Sweden seemed to have a comparable health related quality of life as judged by the preoperative SF-36 scores, although the preoperative differences in the General Health domain suggest the possibility of different comorbidities in the two groups. Both surgical treatments improved HRQoL as described by SF-36 one year postopera-

tively, the more pronounced 1 year improvement seen in the X STOP group might be the result of a different selection of patients and does not allow for anticipated conclusions regarding the new procedure's effectiveness. From a methodological point of view, even though the widespread use of SF-36 as an outcome measure has been a great step towards standardisation of results, and the use of norm-based scoring now allows for international comparisons, caution should still be put in place in the retrospective interpretation of baseline differences and postoperative results across studies.

The low values of correlation coefficients between SF-36 domains and other preoperative clinical variables relating to similar issues might surprise. Especially regarding pain assessment we would have expected stronger correlation between the BP domain and assessment of back and leg pain on the VAS or the consumption of analgesics, such as the one between PF and walking ability. Walsh et al. recently presented data which supported a hypothesis we had previously formulated (Paper I) i.e. that for studies of patients with low back problems, the general SF-36 might be a sufficient measure (Walsh et al 2003). Carefully chosen, variables (like separate VAS assessment for Back and Leg pain intensity) might add useful clinical information (Paper III and IV), which might not be included in the SF-36 scores, without the need for separate condition-specific instruments.

Outcomes can be determined in absolute terms, i.e. prospectively recording the measure at different time points and then calculating the difference, or relatively, asking the patient to judge to what extent his symptoms have decreased or increased after a defined length of time. The latter solution does not imply that the study is retrospective, since the choice of this outcome measure could have been prospectively included in the study design, but leaves us without any data to characterise patient population at the beginning of the study. Findings of papers III, IV and V corroborate the hypothesis that these two approaches do not necessarily lead to the same result. Should we believe our patients, when they say their pain has or has not decreased, or rely solely on the differences between the two VAS measurements? Different types of recall bias have been often advocated to explain differences between the two methods, suggesting that the

absolute prospective one is the correct one (Linton and Melin 1982, Ross 1989, Streiner and Norman 1989).

Even though widely supported, differential scores are less immediate as outcome measures because they introduce arbitrary calculation, which might increase sources of variability and error. Besides, asking the patients if he is feeling better is probably the most ancient of all outcome measures, and a positive answer is the ultimate aim of all our efforts in everyday practice. Most likely absolute and relative measures represent different aspects of the picture, and should be evaluated separately. This is a very controversial issue, but crucial, especially when we try to practice evidence-based musculoskeletal care. If similar methods of assessing pain cannot be assimilated, even less can they be pooled. Not all systematic reviews address this issue properly (Furlan et al. 2001). Cochrane back group has given attention to outcome evaluation in their methodological guideline for reviewer (Van Tulder et al. 1997). In the field of pain, possible solutions have been proposed (Moore et al. 1997). In our opinion, an arbitrary data extraction and pooling of results very often is due to a misinterpretation of the differences between outcome evaluation across trials, and usually constitute the only major flaw in systematic reviews, especially when authors lack direct clinical experience in the field.

A well performed evidence-based critical appraisal must consider the difference between statistical significance and clinical significance. Large samples most likely produce statistically significant results, but the entity of the result can be irrelevant from the patient's perspective. For this reason, Minimally Clinically Important Differences (MCIDs) have been introduced and defined in various settings as well as in LBP (Beaton 2000). These can be useful not only to weigh the results, but also for sample size calculations in the planning phase of a new study. It is important to remember that different clinical settings may imply different MCIDs: for example, MCID for VAS for acute pain (9–13 mm) could be different from that of chronic pain (20 mm), and MCID for patients with little pain at enrolment might differ from that of patients with a high pain level (Todd et al. 1996, Kelly 1999, Bird and Dickson 2001).

No specific study dealing with Minimal Clinically Important Differences (MCIDs) of deviations from SF-36 normative values in LBP seems to exist, although in-depth work has been performed on the other instruments (Bombardier et al. 2001, Hägg et al. 2003). Taking inspiration from other methods of outcome assessment (Roland and Morris, Oswestry, VAS) (Beaton et al. 2001, Bombardier et al. 2001, Hägg et al. 2003, Davidson et

al. 2004), discussion in paper IV and VI speculates that 10–20% of the highest possible value should represent an important clinical difference. Information presented in Tables 16–18 (paper V) facilitates the accomplishment of informed decisions regarding important differences in similar patients: it is important to underline that important differences vary for different outcomes and for different diagnostic groups.

Conclusions

Paper I – The number of proposed outcomes, outcome scores, outcome instruments is incredibly high: no new HRQoL instruments specific for LBP are needed. Instead, effort should be directed toward comparative studies and systematic reviews of the evidence available. Standardisation of outcome evaluation should be a major task for international spine societies.

Paper II – The data collection protocol of the Swedish Spine Register studied can reliably detect postoperative improvements between large groups of patients such as in a Register

Paper III – Measuring pain intensity on a VAS is a useful tool in describing patients scheduled for lumbar spine surgery. Pain intensity measured on the VAS correlates significantly to other indicators of perceived pain, but correlation is not as strong as it could be if they were all measuring the same construct. In the search for a standard in the evaluation of pain as an outcome, the differences between the various methods should be taken into account

Paper IV – HRQoL as measured by SF36 in patients scheduled for lumbar spine surgery showed a pronounced reduction compared to normal and LBP population. Physical domains were more affected than mental domains. The use of norm-based scoring for SF36 can help interpretation and simplify graphic representation of the findings

Paper V – SF-36 outcomes 1-year after surgery for lumbar spine disorders are improved, although

other factors or natural history could contribute to this improvement. Cluster analysis classified the patients into two groups with distinct preoperative patterns but no predictive value as regards postoperative results. The global effect of spine surgery in our sample is quite similar to effect sizes of very successful orthopedic interventions

Paper VI – The patients selected for surgery for spinal stenosis in the US and Sweden seemed to have a comparable health related quality of life as judged by the preoperative SF-36 scores, although the preoperative differences in the General Health domain suggest the possibility of different prevalence of comorbidities in the two groups. Both surgical treatments improved HRQoL as described by SF-36 one year postoperatively. The use of standardised outcome measures allows international comparisons, although caution should be used in the interpretation of differences.

Using analyses of standard National Spine Register data, an increased understanding of some aspects of outcome measurement in patients operated on for lumbar spine problems has been achieved. Reference values and more speculative data (such as effect sizes and MCIDs) have been presented. The *ecological* methodology has been presented and discussed: with a clear respect for its limitations, it could be used to obtain relevant information also in other clinical fields.

Future perspectives

Despite sharing the enormous practical difficulties of running a large multicenter nationwide data collection, Spine and Arthroplasty Registers have little in common. Patient description and outcome evaluation in spine surgery cannot rely on simple categories (*trauma, degenerative or inflammatory disease*) and dichotomous outcomes (*revision*). Thus, registration of spinal surgery is a daunting task compared to registration of joint replacement interventions, and at the same time is far more ambitious because it expands the need for an outcome assessment from limited outcome studies to large databases on a national level. Difficulties of registration and incompleteness of data in the present thesis should be a source of reflection for those dealing with large outcome registries in Sweden and in the rest of the world: it would be an incredible waste of time and resources if similar data – collected with the input of many sources – should not be carefully evaluated (*recycled*). At the same time, the *ecological* methodology proposed can be seen as a new name for an old method, as always: “*nothing new under the sun*” (Eccl 1,9). Besides, in the EBM era, scientific requirements are becoming more strict and analyzing *recycled* data can be a very cumbersome and time-consum-

ing effort. Even bearing all these *caveats* in mind, there might still be room for some future work to do.

Patient-centered evaluation has been a *Copernican revolution* in clinical outcome research but it has not eliminated the need for valid and reliable objective outcomes. In spine surgery, Radio-Stereometric Analysis (RSA) has been proven to fulfil the methodological requirement for a prospective postoperative assessment (Johnsson et al. 1990). The next logical step in the *ecological* methodology would be to *recycle* information from the many Lund RSA studies and cross tabulate it with register information to gain a better understanding of their clinical relevance and predictive value: information obtained could be very useful in planning future RSA studies and interpreting their results.

In the future clinical researchers will need to concentrate on few, valid and recognised standard methods of outcome evaluation. The role and responsibility of every practising physician will remain crucial in the process of translating exoteric acronym such as VAS or MCIDs into useful patient information, possibly the most important skill that makes medicine still an art in the evidence-based era.

Summary

Background

There is no consensus regarding outcomes assessment in spine patients. When using Health-Related Quality of Life (HRQoL) in assessing outcomes of treatment, normative data for different diagnoses are needed to allow comparisons between existing and future studies. The most used generic instrument for HRQoL evaluation in spine pathologies is SF-36. There is no standard definition of important differences in spine surgery patients and no standard reference for minimal clinically important difference.

Objective

The overall aim of this thesis was to increase understanding of some aspects of outcome measurement in patients operated on for lumbar spine problems using existing prospective data available at the spine section of the Orthopedics Department of the University of Lund, and, moreover, to explore potentials and weaknesses in the methodology of retrospectively analysis of prospectively collected observational data.

Methods

A systematic web-search and review of the literature and a retrospective analysis of prospective cohort data collected within the data collection protocol in use by the Swedish Spine Register. Since 1993 all persons undergoing elective lumbar surgery at the spine section of the Orthopedics Department of the University of Lund were included in the prospective registration protocol either in its first version (1993–1997), which was the source of data for paper III, or in its revised version after 1998.

Initial information at baseline (the day before surgery) included age, sex, smoking habits, duration of preoperative back and leg pain in months, duration of preoperative sickleave in months, number of previous operations, patient's working status, diagnostic techniques pre- and post-operatively VAS scores analgesic intake, walking distance. Postoperatively, change in leg and back

pain, respectively, was recorded on a 5-point Likert Scale as compared to preoperative status, patient satisfaction was recorded on a 3-point Likert Scale.

Results

The number of proposed outcomes, outcome scores, outcome instruments is incredibly high: no new HRQoL instruments specific for LBP are needed. The data collection protocol of the Swedish Spine Register studied can reliably detect post-operative improvements between large groups of patients such as in a Register. Pain intensity measured on the VAS correlates significantly to other indicators of perceived pain, but correlation is not as strong as it could be if they were all measuring the same construct. HRQoL as measured by SF-36 in patients scheduled for lumbar spine surgery showed a pronounced reduction compared to normal and LBP population. The use of norm-based scoring for SF-36 can help interpretation and simplify graphic representation of the findings. SF-36 outcomes 1-year after surgery for lumbar spine disorders are improved, although other factors or natural history could contribute to this improvement. The global effect of spine surgery in our sample is quite similar to effect sizes of very successful orthopedic interventions. The use of standardised outcome measures allows international comparisons, although caution should be used in the interpretation of differences.

Conclusions

Without the need of additional expensive data collection and using limited economic resources an increased understanding of some aspects of outcome measurement in patients operated on for lumbar spine problems has been achieved. Reference values and more speculative data (such as effect sizes and MCIDs) have been presented. The *ecological* methodology has been presented and discussed: with a clear respect for its limitations, it could be used to obtain relevant information also in other clinical fields.

Acknowledgements

This thesis would have never been possible if Björn Strömqvist and Bo Jönsson, my co-authors, had not started the prospective registration of data in 1993, if all the colleagues at the spine section had not registered their interventions, if all the patients had not completed their forms, if Kaj Knutson had not prepared the computerised database and Lena Oreby had not entered the data in it.

I would have never come to Lund if I had not met Lars Lidgren many years ago and if he had not been such a tremendous fisher of men. I would have never met Lars if Francesco Greco, my former director, had not allowed me to host him for a scientific meeting at the Department of Orthopaedics of University of Ancona, where I begun to follow the track of my late father, Silvio Zanoli, former director, and his legacy as an orthopaedic surgeon. What I never could learn from him, I derived from the marvellous tutors and residents of that time.

This work started after my activity at the Spine Section in Lund as a Marie Curie post-doctoral fellow on RSA, and would have never come to life without the unrestricted support from Björn Strömqvist, my Mentor and Friend, who believed in its vague conception at the very beginning and precisely corrected until the last figure at the very end, keeping my colourful wording within boundaries of scientific language and burying with a laugh stubborn peer-reviewers, always giving me the impression that I was as young, motivated and enthusiastic as he actually is.

All the discussions and all the things I learned from persons in Lund, in Italy and around the world along the years (e.g. co-authors, opponents, authors of previous theses on similar subjects, relatives, friends, teachers, researchers, nurses, physiotherapists, patients or simple acquaintances) are somehow reflected in this thesis: although it would be unfair to try to mention all, many will recognise the seeds of their contribution here or there. Emilio Romanini, lifelong friend and companion, should rather look for a grown-up tree, since the intuitions and discussions I had with him and Roberto Padua

10 years ago when GLOBE and *e_Musk* started are still fruitful to date: if a thesis could have more than an author, they would probably be my co-authors once more.

It would have not been possible for me to continue these studies without the understanding of all the institutions I worked for in Italy and at an international level. In particular Giancarlo Traina, my present director at the Department of Orthopaedics of University of Ferrara, even encouraged me to proceed, Leo Massari and all the colleagues and residents at the department had to bear with my frequent disappearances. Last but not least, my patients and students, had to adjust to a very irregular schedule.

Coming to Sweden would have been much more difficult without SAS direct flight Bologna-Copenhagen and the Öresund bridge, without the possibility to stay at the Anhörighotell, and without Lena Oreby and Gun-Britt Nyberg settling all these bureaucratic issues for me. Living in Sweden would have been much harder without Gloria's serie A matches, without Torsdagsfika in the company of the friends at the Laboratory of Biomechanics and most of all without Ninni and Björn taking care of my Swedish language and culture between a warm meal (*tack för senast!*) a tennis match and a North Sea sauna.

Life is wonderful, even besides a PhD, and full of feelings and persons which will remain private as they do not belong to this thesis, but make it worthwhile: my family was always there to remind it to me, as they do now. My best acknowledgement will be to stop writing and go back to them.

Funding: Swedish National Board of Health and Welfare, Medical Faculty, University of Lund, Stiftelsen för Bistånd åt Rörelsehindrade i Skåne and Vetenskapsrådet (09509) sustained infrastructural costs of the project throughout the years and reimbursement of direct expenses. My mother, Amelia Guglielmi, University and Hospital of Ferrara, private companies and patients provided the rest.

References

- Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study, Part III. 1-year outcomes of surgical and nonsurgical management of lumbar spinal stenosis. *Spine* 1996; 21: 1787-94; discussion 94-5.
- Atlas SJ, Keller RB, Robson D, et al. Surgical and nonsurgical management of lumbar spinal stenosis. *Spine* 2000; 25: 556-62.
- Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, Strand V, Shea B. Looking for important change/differences in studies of responsiveness. OMER-ACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference. *J Rheumatol* 2001; 28(2): 400-5.
- Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine* 2000; 25(24): 3192-9.
- Bombardier C, Hayden J, Beaton DE. Minimal clinically important difference. Low back pain: outcome measures. *J Rheumatol* 2001; 28(2): 431-8.
- Browaeys MJ. Complexity of epistemology: Theory of knowledge or philosophy of science? Fourth Annual Meeting of the European Chaos and Complexity in Organisations Network (ECCON), Driebergen, 2004 URL: <http://www.chaosforum.com/nieuws/CofE.pdf> (visited 2005 March 18th)
- Brox JI, Sorensen R, Friis A, Nygaard O, Indahl A, Keller A, Ingebrigtsen T, Eriksen HR, Holm I, Koller AK, Riise R, Reikeras O. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine* 2003; 28(17): 1913-21.
- Chatman AB, Hyams SP, Neel JM, Binkley JM, Stratford PW, Schomberg A, Stabler M. The Patient-Specific Functional Scale: measurement properties in patients with knee dysfunction. *Phys Ther* 1997; 77(8): 820-9.
- Conover WJ. Practical Nonparametric Statistics. John Wiley and Sons, New York, 1980; cit. in Bland M, An Introduction to Medical Statistics. Oxford University Press, 2000
- Davidson M, Keating JL, Eyres S. A low back-specific version of the SF-36 Physical Functioning scale. *Spine* 2004; 29(5): 586-94.
- Deyo R, Cherkn D, Loeser J, Bigos S, Ciol M. Morbidity and mortality in association with operations on the lumbar spine. The influence of age, diagnosis, and procedure. *J Bone Joint Surg* 1992; 74-A: 536-43.
- Deyo RA, Nachemson A, Mirza SK. Spinal-fusion surgery - the case for restraint. *N Engl J Med* 2004; 350(7): 722-6.
- Draper SW. (2005 February 2nd) The Hawthorne effect [WWW document]. URL: <http://www.psy.gla.ac.uk/~steve/hawth.html> (visited 2005 March 18th)
- Dunbar MJ, 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.
- Dunbar MJ, Robertsson O, Ryd L. What's all that noise? The effect of co-morbidity on health outcome questionnaire results after knee arthroplasty. *Acta Orthop Scand* 2004; 75(2): 119-2.
- Dunbar MJ. Subjective outcomes after knee arthroplasty. *Acta Orthop Scand* 2001; 72(Suppl 301):1-63
- Evans JH, Kagan A 2d. The development of a functional rating scale to measure the treatment outcome of chronic spinal patients. *Spine* 1986; 11(3): 277-81.
- Fairbank J. Point of view on: Validation of the Turkish version of the Oswestry Disability Index for patients with low back pain. *Spine* 2004; 29(5): 585.
- Fairbank JC, Frost H, Wilson-MacDonald J, Yu LM, Barker K, Collins R. The MRC Spine Stabilisation Trial. A randomised controlled trial to compare surgical stabilisation of the lumbar spine versus an intensive rehabilitation programme on outcome in patients with chronic low back pain. In Abstract Book (ISSLS), Spine Week Porto 2004.
- Fanciullo GJ, Hanscom B, Weinstein JN, Chawarski MC, Jamison RN, Baird JC. Cluster analysis classification of SF-36 profiles for patients with spinal pain. *Spine* 2003; 28(19): 2276-82.
- Fritzell P, Hagg O, Wessberg P, Nordwall A. 2001 Volvo Award Winner in Clinical Studies: Lumbar fusion versus nonsurgical treatment for chronic low back pain: a multicenter randomized controlled trial from the Swedish Lumbar Spine Study Group. *Spine* 2001; 26(23): 2521-32; discussion 32-4.
- Furlan AD, Clarke J, Esmail R, Sinclair S, Irvin E, Bombardier C. A critical review of reviews on the treatment of chronic low back pain. *Spine* 2001; 26(7): E155-62.
- Gartsman GM, Khan M, Hammerman SM. Arthroscopic repair of full-thickness tears of the rotator cuff. *J Bone Joint Surg Am* 1998; 80(6): 832-40.
- Gibson JN, Grant IC, Waddell G. The Cochrane review of surgery for lumbar disc prolapse and degenerative lumbar spondylosis. *Spine* 1999; 24(17): 1820-32.
- Gibson JN, Waddell G, Grant IC. Surgery for degenerative lumbar spondylosis. *Cochrane Database Syst Rev* 2000; (3): CD001352.
- 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.
- Hee HT, III Whitecloud TS 3rd, Myers L, Roesch W, Ricciardi JE. Do worker's compensation patients with neck pain have lower SF-36 scores? *Eur Spine J* 2002; 11(4): 375-81.
- Herno A, Airaksinen O, Saari T. Long-term results of surgical treatment of lumbar spinal stenosis. *Spine* 1993; 18: 1471-4.
- Huskisson EC. Measurement of pain. *Lancet*. 1974; 2(7889): 1127-31.

- Hägg O. Measurement and prediction of outcome. Application in fusion surgery for chronic Low-Back Pain. Göteborg University, Thesis, Göteborg, Sweden 2002
- Hägg O, Fritzell P, Nordwall A; Swedish Lumbar Spine Study Group. The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 2003; 12(1): 12-20; discussion 21.
- Jamison, RN, Rock DL, Parris WCV. Empirically derived Symptom Checklist 90 subgroups of chronic pain patients: a cluster analysis. *J Behav Med* 1987; 11: 147.
- Jansson KA, Nemeth G, Granath F, Blomqvist P. Surgery for herniation of a lumbar disc in Sweden between 1987 and 1999. An analysis of 27,576 operations. *J Bone Joint Surg Br* 2004; 86(6): 841-7.
- Jansson K-Å, Blomqvist P, Granath F, Németh G. Spinal stenosis surgery in Sweden 1987-1999. *Eur Spin J* 2003; 12: 535-41.
- Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. *Pain* 1986; 27: 117-26.
- Johnsson KE, Rosen I, Uden A. The natural course of lumbar spinal stenosis. *Clin Orthop* 1992; (279): 82-6.
- Johnsson KE, Uden A, Rosen I. The effect of decompression on the natural course of spinal stenosis. A comparison of surgically treated and untreated patients. *Spine* 1991; 16: 615-9.
- Johnsson R, Selvik G, Strömquist B, Sundén G. Mobility of the lower lumbar spine after posterolateral fusion determined by roentgen stereophotogrammetric analysis. *Spine* 1990; 15(5): 347-50.
- Keller A, Brox JI, Gundersen R, Holm I, Friis A, Reikeras O. Trunk muscle strength, cross-sectional area, and density in patients with chronic low back pain randomized to lumbar fusion or cognitive intervention and exercises. *Spine* 2004; 29(1): 3-8.
- Keller RB, Atlas SJ, Soule DN, Singer DE, Deyo RA. Relationship between rates and outcomes of operative treatment for lumbar disc herniation and spinal stenosis. *J Bone Joint Surg Am* 1999; 81(6): 752-62.
- Kelly AM. Does the clinically significant difference in visual analog scale pain scores vary with gender, age, or cause of pain? *Acad Emerg Med* 1998; 5(11): 1086-90.
- Linton SJ, Melin L. The accuracy of remembering chronic pain. *Pain* 1982; 13(3): 281-5.
- 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.
- McKinnon ME, Vickers MR, Ruddock VM, Townsend J, Meade TW. Community studies of the health service implications of low back pain. *Spine* 1997; 22(18): 2161-6.
- Mele A, Bianco E, Torre M, Wenzel V, Romanini E, Padua R, Zanolli G. Revisione sistematica sulle protesi d'anca: affidabilità dell'impianto. Istituto Superiore di Sanità - PNLG, Zadig, Milano 2004
- Moore A, Moore O, McQuay H, Gavaghan D. Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics: use of pain intensity and visual analog scales. *Pain* 1997; 69(3): 311-5.
- Morin E. La méthode: La connaissance de la connaissance. Paris: Seuil 1986
- Ostelo RW, de Vet HC, Knol DL, van den Brandt PA. 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.
- Ostendorf M, van Stel HF, Buskens E, Schrijvers AJ, Marting LN, Verbout AJ, Dhert WJ. Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. *J Bone Joint Surg Br* 2004; 86(6): 801-8.
- Robertsson O. The Swedish knee arthroplasty register. Validity and Outcome. University of Lund, Lund. Thesis 2000.
- Roos E. Knee Injury and Knee Osteoarthritis -Development, evaluation and clinical application of patient-relevant questionnaires. University of Lund, Lund. Thesis 1999.
- Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)-development of a self-administered outcome measure. *J Orthop Sports Phys Ther* 1998; 28(2): 88-96.
- Roos H, Roos E, Ryd L. On the art of measuring. *Acta Orthop Scand* 1997; 68(1): 3-5.
- Rosenthal GE, Harper DL, Quinn LM, Cooper GS. Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals. Results of a regional study. *JAMA* 1997; 278(6): 485-90.
- Ross M. Relation of implicit theories to the construction of personal histories. *Psych Rev* 1989; 96: 341-57.
- Schofferman J. Long-term opioid analgesic therapy for severe refractory lumbar spine pain. *Clin J Pain* 1999; 15(2): 136-40.
- Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 2nd Ed. New York: Oxford University Press; 1989.
- Strombeck B, Ekdahl C, Manthorpe R, Wikstrom 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.
- 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.
- Strömquist B, Jönsson B. Computerized follow-up after surgery for degenerative lumbar spine diseases. *Acta Orthop Scand* 1993; (Suppl 251): 138-42.
- Stucki G, Liang MH, Phillips C, Katz JN. 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 JE, Hälsoenkät SF-36. *Svensk Manual och Tolkningsguide (SF-36 Health Survey. Swedish Manual and Interpretation Guide)*, Sahlgrenska University Hospital, Göteborg, 1994.
- Tidemark J, Bergstrom G, Svensson O, Tornkvist H, Ponzer S. Responsiveness of the EuroQol (EQ 5-D) and the SF-36 in elderly patients with displaced femoral neck fractures. *Qual Life Res* 2003; 12(8): 1069-79.
- Todd KH, Funk KG, Funk JP, Bonacci R. Clinical significance of reported changes in pain severity. *Ann Emerg Med* 1996; 27(4): 485-9.

- Turk DC, Rudy TE. Toward an empirically derived taxonomy of chronic pain patients: integration of psychological assessment data. *J Consult Clin Psych* 1988; 56: 233-8.
- Walsh TL, Hanscom B, Lurie JD, Weinstein JN. 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.
- Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Qual Life Outcomes* 2003; 1(1): 4.
- van Tulder MW, Assendelft WJ, Koes BW, Bouter LM. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. *Spine* 1997; 22(20): 2323-30.
- Ware JE Jr, Bjorner JB, 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* 2000; 38(9 Suppl): II73-82.
- Ware JE Jr, Sherbourne CD. The MOS 36-item Short-Form health survey (SF-36). 1. Conceptual framework and item selection. *Med Care* 1992; 30: 473-83.
- Ware JE Jr. SF-36 health survey update. *Spine* 2000; 25(24): 3130-9.
- Ware JE, Snow K, Kosinski M, et al. SF-36 Health Survey: Manual and Interpretation Guide. Boston, MA: The Health Institute; 1993.
- Vogt MT, Hanscom B, Lauerman WC, Kang JD. Influence of smoking on the health status of spinal patients: the National Spine Network database. *Spine* 2002; 27(3): 313-9.
- Von Korff M, Jensen MP, Karoly P. Assessing global pain severity by self-report in clinical and health services research. *Spine* 2000;25(24): 3140-51.
- Zanoli G, Romanini E, Padua R, Traina GC, Massari L. EBM in musculoskeletal diseases: where are we? *Acta Orthop Scand* 2002; 73(Suppl 305): 4-7.
- Zanoli G, Strömqvist B, Jönsson B. SF-36 for outcomes assessment of spine surgery. A prospective consecutive study of 451 operated patients. *Proceedings, Spineweek Oporto*, 2004.
- Zucherman JF, Hsu KY, Hartjen CA, Mehalic TF, Implicito DA, Martin MJ, Johnson DR 2nd, Skidmore GA, Vessa PP, Dwyer JW, Puccio S, Cauthen JC, Ozuna RM. A prospective randomized multi-center study for the treatment of lumbar spinal stenosis with the X STOP interspinous implant: 1-year results. *Eur Spine J* 2004; 13(1): 22-31.

Summary in Swedish – Svensk sammanfattning

Hur man värderar resultat av kirurgisk behandling av ländryggens degenerativa åkommor är ett stort, komplext och delvis kontroversiellt ämne. Denna avhandling studerar resultatvärdering med olika metoder och instrument, huvudsakligen baserat på resultat från det nationella svenska ländryggskirurgiregistret.

Arbete I. Inventering av befintliga instrument i litteraturen avseende hälsorelaterad livskvalitet vid ryggsjukdom. Nittiotvå frågeinstrument kunde identifieras, mer än hälften publicerade för första gången under de senaste 10 åren. Internationell enighet om lämpliga utvärderingsinstrument behövs, inte ytterligare nykonstruerade utvärderingsinstrument.

Arbete II. Validering av data från svenska ryggregisterprotokollet. Test-retest analys av 119 patienter, 63 av dessa fyllde i det preoperativa protokollet två gånger och 59 andra patienter fyllde i det ett år postoperativa protokollet två gånger. Reliabilitetsanalyser gjordes relaterade till tidsintervallet mellan första och andra ifyllande. Reliabiliteten minskade om intervallet översteg 3 veckor i den preoperativa gruppen och 9 veckor i den postoperativa gruppen. Protokollet har förmåga att dokumentera resultatförändringar efter kirurgi.

Arbete III. Smärtmätning på VAS-skala före och efter kirurgisk behandling och korrelation mellan olika former att mäta smärtlindring utfördes på 755 konsekutiva patienter. Smärtprofilerna klart särskilda mellan de olika sjukdomsdiagnoserna, snabb regress postoperativt för medelvärden för VAS-smärta såväl avseende rygg- som bensmärta. Signifikanta korrelationer mellan patienttillfredsställelse av kirurgi och andra "smärtfacit" såsom upplevd förändring i smärta, VAS-smärta vid uppföljning samt analgeticakonsumtion.

Arbete IV. Livskvalitetsmätning med SF-36 på 451 patienter planerade för ländryggskirurgisk åtgärd. Normativa data för dessa sjukdomsgrupper etablerades och visade mycket låga värden på SF-36 profilerna jämfört med friska människor i samma ålderskategori, inte bara beträffande

fysiska utan även psykiska domäner. Likaså visade den aktuella sjukdomsgruppen sämre livskvalitet än patienter med fibromyalgi, reumatoid artrit, migrän och ytterligare några andra sjukdomar. Ett stort basmaterial som norm för framtida studier har presenterats.

Arbete V. SF-36 såsom resultatutvärderande instrument efter ryggkirurgi evaluerades. Trehundrafemtioen patienter fyllde i SF-36 formuläret preoperativt och 1 år postoperativt. Korrelation mellan SF-36 domäner och andra resultatparametrar studerades. Stark korrelation mellan minskad kroppslig smärta i SF-36 och minskad smärta på VAS-skalan sågs, likaså korrelerade fysisk funktion starkt till förbättring i gångfunktionen. En clusteranalys kunde skilja patienter med förmåga till emotionell adaption från patienter med mera dysfunktionellt beteende, förstnämnda gruppen med låga fysiska och ganska normala mentala domäner, sistnämnda gruppen med uttalad reduktion i båda aspekterna. Postoperativt förbättrades båda patientkategorierna i stort sett i samma omfattning.

Arbete VI. Möjligheten att jämföra SF-36 profiler i två olika patientmaterial i olika länder med samma diagnos studerades. Nittio patienter i en FDA-studie av amerikanska patienter opererade för spinal stenosis med en ny operationsteknik, indirekt dekompression i lokalanestesi, jämfördes med 90 ålders- och könsmatchade patienter, opererade för spinal stenosis vid ortopediska kliniken i Lund på konventionellt vis med dekompressionsoperation. Livskvalitetsprofilerna matchades till nationsspecifika ålders- och könsnormerade data. Patientselektionen inte likartad i grupperna men möjligheten att med gemensamma resultatparametrar kunna jämföra olika patientmaterial påvisades och kommer att kunna användas i framtiden.

Sammanfattningsvis har denna avhandling skapat referensbanker för framtida studier av ländryggskirurgi och presenterat en analys av olika resultatutvärderingsparametrar och deras inbördes relation.

Appendix

Forms

1993–1997 (Paper III)



SPINAL SURGERY Dpt _____ No: _____

Radiology

Myelography	☐ Yes	☐ No
CT	☐ Yes	☐ No
Myelo-CT	☐ Yes	☐ No
MRI	☐ Yes	☐ No
Diagnostic block	☐ Yes	☐ No

Pain drawing

Pain drawing (1–4) _____

Diagnosis

- ☐ Disc herniation
☐ Central spinal stenosis
☐ Lateral spinal stenosis
☐ Spondylolysis/olisthesis
☐ Segmental pain
☐ Other.....

Operation

- ☐ Open discectomy
☐ Microscopic discectomy
☐ Percutaneous nucleotomy
☐ Decompression
☐ Decompression + posterior fusion, uninstrumented
☐ Decompression + posterior fusion, instrumented
☐ Posterior fusion, uninstrumented
☐ Posterior fusion, instrumented
☐ Anterior fusion, uninstrumented
☐ Anterior fusion, instrumented
☐ Other.....

Type of implant.....

Level operated on (from – to)

(i.e. nerve root: L4 – L4)

Proximal

(i.e. disc: L5 – S1)

Distal

Side

☐ Bilateral

☐ Right

☐ Left

Antibiotic prophylaxis

☐ Yes

☐ No

Patient ID _____

Strömqvist/Jönsson

Hospitalization

Date (yy-mm-dd)

Admitted

Date of operation

Discharged

Surgeon (initials)

Postoperative complication

☐ No

☐ Yes, type

Reoperation

A new operation on the same side and level or an operation for a postoperative complication following the primary procedure is regarded as a reoperation. Only indication, date and type of procedure is recorded for a reoperation.

Indication for reoperation

Date of reoperation

(yy-mm-dd)

Reoperation

- ☐ Discectomy
☐ Decompression
☐ Decompression + posterior fusion
☐ Posterior fusion
☐ Anterior fusion
☐ Reoperation for complication
☐ Other.....

Type of implant.....

Reviewer

Reviewer, initials

1998–2003 (Paper II, IV, V, and VI)

**SPINAL SURGERY**

No: _____

Operation

This is a

- ☐ 1. Primary operation
- ☐ 2. Operation within 12 months of a previous spinal op.
- ☐ 3. Operation >12 months from a previous spinal op.

Reoperation of a postoperative complication during ongoing hospital stay is recorded on the primary form

Pain drawing

Pain drawing (1–4)

1= Nonorganic, 2= Possibly nonorganic, ☐
 3= Possibly organic, 4=Organic ☐

Diagnosis

- ☐ 1. Disc herniation
- ☐ 2. Central spinal stenosis
- ☐ 3. Lateral spinal stenosis
- ☐ 4. Spondylolysis/olisthesis
- ☐ 5. Segmental pain
- ☐ Other.....

Type of operation

- ☐ 1. Open discectomy
- ☐ 2. Microscopic discectomy
- ☐ 3. Percutaneous nucleotomy
- ☐ 4. Decompression
- ☐ 5. Decompression + posterior fusion, uninstrumented
- ☐ 6. Decompression + posterior fusion, instrumented
- ☐ 7. Posterior fusion, uninstrumented
- ☐ 8. Posterior fusion, instrumented
- ☐ 9. Anterior fusion, uninstrumented
- ☐ 10. Anterior fusion, instrumented
- ☐ Other.....

Type of implant.....

Level operated on (from – to)

(i.e. nerve root: L4 – L4)

Proximal ☐

(i.e. disc: L5 – S1)

Distal ☐

Side

☐ Bilateral☐ Right☐ Left

Antibiotic prophylaxis

☐ Yes☐ No

Patient ID

Strömqvist/Jönsson

Hospitalization

Date (yy-mm-dd)

Admitted

Date of operation

Discharged

Surgeon (initials)

Postoperative compl. during stay

Postoperative complications during hospital stay

- | | | |
|----------------------------|------------------------------|-----------------------------|
| 1. Death | ☐ Yes | ☐ No |
| 2. Thrombosis | ☐ Yes | ☐ No |
| 3. Pulmonary embolism | ☐ Yes | ☐ No |
| 4. Urinary tract infection | ☐ Yes | ☐ No |
| 5. Bleeding/hemtoma | ☐ Yes | ☐ No |
| 6. Wound infection | ☐ Yes | ☐ No |
| 7. Nerve root damage | ☐ Yes | ☐ No |
| 8. Cauda equina syndrome | ☐ Yes | ☐ No |
| 9. Dural lesion | ☐ Yes | ☐ No |
| 10. Malplaced hardware | ☐ Yes | ☐ No |
| 11. Other | | |

Reoperation during the hospital stay

Surgical intervention because of complication

☐ Yes☐ No

Type of procedure

Reviewer

Department

Reviewer, initials

Literary and figurative references

Einstein said he could see so far because he stood on the shoulders of giants. Unfortunately, I wear glasses, so I did not achieve quite as much. However, my giants taught me to have a close look as well, and enjoy a good book and many forms of art. My giants, my roots, my many "families": Zanolli, Guglielmi, Chemello, Andersson, Arend-Heidbrincks, Strömqvist ...

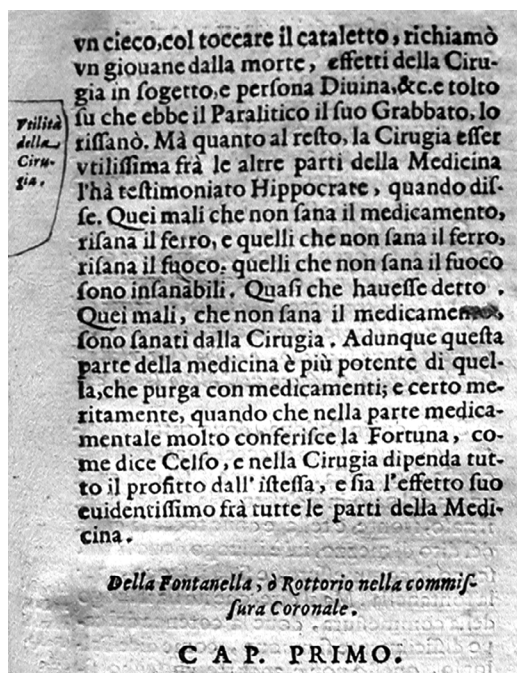


Figure 13. Girolamo Fabritio d'Aquapendente (1533–1619) *L'opere cirurgiche*. Padua 1671 Second part: On surgical operations, Prologue p. 189

Und plötzlich in diesem mühsamen Nirgends, plötzlich / die unsägliche Stelle, wo sich das reine Zuwenig / unbegreiflich verwandelt -, umspringt / in jenes leere Zuviel. / Wo die vielstellige Rechnung / zahlenlos aufgeht (And suddenly in this laborious nowhere, suddenly / the unsayable place, where the pure too-little / inexplicably changes - , leaps / into that empty too-much. / Where the many-numbered calculation / numberlessly resolves).

RM Rilke Duineser Elegien – Die fünfte Elegie (1923)

E andando nel sole che abbaglia / sentire con triste meraviglia / com'è tutta la vita e il suo travaglio / in questo seguire una muraglia / che ha in cima cocci aguzzi di bottiglia (And walking in the dazzling sunlight / to feel with sad wonder / how all life and its labour / is in this tracing along a wall / with jagged bits of bottle on top).

E Montale, *Ossi di Seppia* (1925)

The wounded surgeon plies the steel / That questions the distempered part; / Beneath the bleeding hands we feel / The sharp compassion of the healer's art / Resolving the enigma of the fever chart.

TS Eliot, *For Quartets* - East Coker (1940)



Figure 14. Pierre Bonnard (1867–1947), *Le Jardin dans le Var* (Garden in Southern France, Var) 1914, oil on canvas 51.0 (h) x 57.0 (w) cm. Villa Flora, Winterthur.

Bonnard worked exclusively from memory. While he observed everything, with his camera, or in his diary or on scraps of paper that he carried with him, when it came to painting even his portraits were done in the absence of the model. Defending this work method, Bonnard explained that having the actual subject in front of him would distract him from his work.

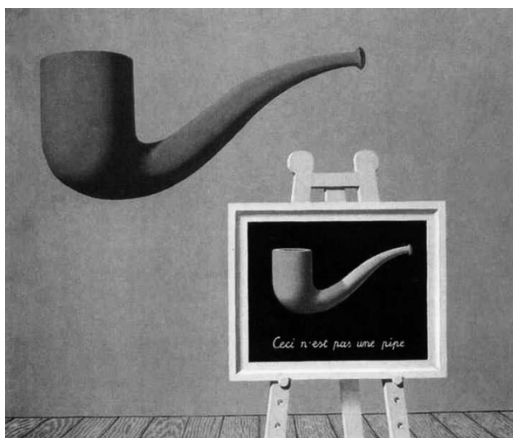


Figure 15. René Magritte (1898–1967), *Les deux mystères* (The two mysteries) 1966, oil on canvas, 60 x 80 cm, private collection, London.

With his incredible skill at painting realistic objects and figures Magritte decided to make each of his painting a visual poem. He said: *"My painting is visible images which conceal nothing; they evoke mystery and, indeed, when one sees one of my pictures, one asks oneself this simple question, 'What does that mean?'. It does not mean anything, because mystery means nothing either, it is unknowable."* This is not a pipe, but which one?



Figure 16. Paul Klee (1878–1940), *Erinnerung an einen Garten* (Remembrance of a Garden), 1914, watercolor on linen paper mounted on cardboard, 25.2 x 21.5 cm. Kunstsammlung Nordrhein-Westfalen, Düsseldorf.

In 1914 Klee went to Tunisia (where maybe this garden was) had already been to Italy (where they say he was impressed by the lights of mosaics). Someone sees a clear link between some of Klee's plant motifs and the images of plankton, diatoms, seeds, and micro-organisms that German scientific photographers were making at the same time. What was that garden really like? How is it today? How much has it changed?

Final notes

Most people will find data presented in this thesis useless as it does not come from an experimental design or a RCT: in fact we could never find a second Italian orthopaedic surgeon to take a PhD in Sweden so we were not able to randomise.

Despite claiming for some *ecological* descent, this thesis will humbly contribute to the global deforestation. If you don't like or you are not going to use it in any conventional sense, please throw it away in the dedicated paper *recycling* bins.