Peer reviewed article

# Symptom assessment in lumbar stenosis/ spondylolysis – patient questionnaire versus physician chart

*Joerg Ferdinand Huber*[a], *Essam Dabis*[a], *Juerg Huesler*[b], *Georg Bernhard Ruflin*[a]

[a] Orthopedic Clinic, Kantonsspital AG, Aarau, Switzerland
[b] Institute of Mathematical Statistics and Insurance Medicine, University of Berne, Switzerland

## Summary

*Background:* The precision of symptom assessment and documentation in daily routine is unknown. Differences in hip studies have been observed between the charts and the patient questionnaires. This was the reason to perform this study and to compare the charts and the patient questionnaires.

*Methods:* This was a retrospective study of physician versus patient symptom ratings in spondylolysis/spinal stenosis patients before and after lumbar spinal surgery. Symptoms were assessed just prior to surgery and 3 and 12 months after surgery, by patients using self-administered validated questionnaire (NASS lumbar element) and by physicians recording notes in patient charts. Presence or absence of symptoms was analysed for back/buttock pain, leg pain, and neurological signs. Agreement between patient and physician rating was scored as total (agreement in all 3 categories), partial (agreement in 1 or 2 categories) or none, at each time point.

*Results:* A total of 44 patients completed the study; average age was 63 years; 68% were women. Symptoms in all 3 symptom categories were reduced significantly 3 and 12 months after surgery, especially leg pain and neurological symptoms. Total agreement between patient and physician ratings of symptoms was substantially lower after surgery (22–28%) than before (50%). This was mainly attributable to physicians reporting less leg pain and neurological signs after surgery than patients. Total agreement was lower for older patients.

*Conclusions:* The agreement of main symptoms between the charts and the patient questionnaires seems to be low and may be ameliorated with more time and more structured interviews to ensure that the physician's patient chart more closely reflects the patient's self-rating, than was the case in this study.

*Key words: assessment; agreement; lumbar surgery; pain; NASS; questionnaires*

## Introduction

Two different ways exist to assess and document symptoms; in a direct way with self administered questionnaires and an indirect way by the physician taking the patient's history and documenting the information in charts.

Validated questionnaires are increasingly used in clinical research for quantifying healthcare outcomes. Such questionnaires are filled out by the patient and allow a direct assessment and documentation of symptoms. These tools have become more widespread in the last twenty years, especially for clinical studies to measure the effects of medical or surgical interventions [2, 4, 7].

In clinical practice, however, assessment of outcome is still largely based on the patient's medical chart, which is filled out by the physician. Taking the patient's history, assessing symptoms and documenting and filing the findings is a demanding task. With experience, the physician becomes more skilled and learns to elicit and document information more efficiently. However this process is still an indirect documentation of symptoms: the physician notes in the charts what he or she thinks to have understood from the patient.

How accurate and reliable are the assessment and documentation of symptoms in routine practice by the physician? In contrast to the extensive literature on validated questionnaires, surprisingly few studies have compared patient- and physician-derived data [5, 10, 12]. In the present study we wanted to compare the symptoms before and after lumbar surgery as recorded by patients (self-administered questionnaire) and physicians (patient chart).

The statistical analysis of this study was financially supported by the Fonds "Wissenschaft und Weiterbildung" of the Kantonsspital.

No conflict of interest.

Since 1999, the symptoms and disabilities of degenerative lumbar spine disease patients in our clinic have been routinely assessed using self-administered generic (SF-36) and specific (North American Spine Society (NASS) lumbar) questionnaires. The NASS lumbar questionnaire [1] quantifies symptoms and disabilities in patients with degenerative lumbar spine disorders. The main dimensions are back-buttock pain, leg pain, and neurological symptoms such as weakness/tingling/numbness and disabilities, based on the Oswestry Disability Index. These symptoms, especially pain, have been shown to be a highly responsive parameter in patients with degenerative spinal disorders [13]. The questionnaire has been validated and tested in the German version [8, 9, 11]. The questionnaire is filled out by the patient without help. In addition, all patients' symptoms are recorded by the physician on patient charts (hand written at admission, dictated following subsequent consultations). The aim of the study was to assess how closely the validated questionnaire data generated by the patient corresponded to the clinical notes made by the physician.

## Patients and methods

### Study design

This was a retrospective, non-interventional, monocentric study. This design was selected intentionally to reduce potential bias; in a prospective situation the physicians generating the chart data could have been influenced by the knowledge of the study goals. Approval by institutional review board and informed consent were not sought because the study constituted a medical quality audit: the data, which were anonymised for analysis, were all collected as part of routine care and did not involve any experimental intervention.

### Patients

The patient population consisted of successive admissions to the clinic for lumbar spinal surgery between 1999 and 2002. Inclusion criteria were a positive history of at least six months with lumbar pain and corresponding radiological findings (X-ray, MRI) of lumbar spine pathology (spondylolysis/spinal stenosis) not responding to conservative treatment. Exclusion criteria were trauma, neoplasia, infection, history of failed back surgery, cognitive disorders, inadequate understanding of German, or missing patient chart records. After surgery (decompression alone or with additional instrumented fusion) all patients were discharged home with instructions for isometric exercises, and returned for out-patient evaluation 3 and 12 months post-surgery.

### Data collection

Questionnaire: All patients are routinely assessed before and after treatment with validated generic and specific questionnaires sent by post for self-completion. The generic questionnaire used is the Short Form 36 (SF-36), and the specific questionnaire for patients with lumbar spine diseases is the lumbar spine outcome assessment instrument of the North American Spine Society (NASS lumbar element), both in the validated German versions [1, 8, 9, 11].

Patient chart: Patient symptoms are routinely documented pre- and post-surgery in standard patient charts. Before surgery, patients are assessed by the surgeon/resident, and symptoms are noted by hand on the appropriate form. After post-surgical consultation in the outpatient clinic, clinical status and symptoms are dictated by the consultant and subsequently typewritten. In this study, all outpatient consultations were performed by the same two surgeons who operated on the patients. The operating surgeons were blinded to the answers in the postal questionnaires sent to the patients during the study.

### Data analysis

Patient questionnaire: The NASS data were collated separately for the three pivotal categories: back-buttock pain (questions 1, 2, 5, 6), leg pain (questions 3, 4, 7, 8), and neurological/disability symptoms (questions 9–16). For the purposes of comparison with the patient chart data, patient responses were transformed to nominal scores, i.e., present or absent. The responses to the question 1 or 2 (= "never" and "seldom") were both scored as "absent", responses 3 to 6 (= "occasionally" to "always") as "present".

Physician chart: The physician charts were analysed by two independent coworkers together for the presence of back-buttock pain, leg pain, and neurological symptoms. Each of the three categories was scored as "present" in a given patient if there was any information interpreted as indicating the presence of that category. All charts were peer-reviewed to check concordance of the grading.

Questionnaire versus chart: Agreement between questionnaire and patient chart for the three symptom categories (back-buttock pain, leg pain, and neurological signs/disability) was calculated separately for pre-surgery baseline and 3 and 12 months post-surgery, as follows: agreement was reported as "total" if questionnaire and patient chart had the same score (i.e., present or absent) for all 3 symptoms, as "partial" if 1 or 2 symptoms had the same score, and as «none« if scores differed between questionnaire and chart for all three symptoms.

### Statistical methods

This study was considered as a pilot study since no preliminary data were available. The value of this study is thus limited because of the small sample size. The analysis of the data is based on statistical descriptive methods and Cohen's kappa for the agreement of two judgments. Confidence intervals are determined for the discussion of the relevance of the statistical results.

## Results

Of 45 eligible patients admitted for lumbar spinal surgery between 1999 and 2002, one was excluded due to inadequate case history documentation; 44 patients were included in the study. The average age of patients was 62.7 years (range 32 to 80, median 66 years), and 30 (68%) were women. The indication for surgery was degenerative lumbar spinal stenosis in 37 (84%), spondylolisthesis in 7 (16%). Surgical treatment was decompression of the affected segments in 38 (86%), decompression and stabilization in 4 (9%), stabilization only in 4 (9%). All 44 patients completed the study.

The individual data on agreement between questionnaire and patient chart are summarised in table 1. Concordance between physician chart and patient questionnaire was calculated from these data in terms of total, partial and no agreement, as described in the Methods section. The results are summarised in figure 1. Prior to surgery, 47.7% of patients had total agreement with physician ratings and 52.2% had partial agreement. At 3 and 12 months post-surgery, the rate of total agreement was approximately halved (to 22.7% at 3 months and 22.7% at 12 months), and the rate of partial agreement correspondingly increased (68.2% and 70.5% respectively). The rate of no agreement between patient and physician rating was zero before and 9.1% and 6.8% 3 months and 12 months after surgery, respectively.

The Cohen's kappa was determined, a measure for the (total) agreement which is 0.05 at baseline, because most ratings were positive in each of the three categories. At 3 months we observed a kappa = 0.10 and a kappa = 0.11 at 12 months after surgery.

Deciding that it was difficult to differentiate back and leg pain, the two categories of back and

**Table 1**

Individual concordance data between physician rating (by chart) (columns) and patient rating (by questionnaire) (rows). "+" = present, "–" = absent. The first position corresponds to back- and buttock pain, the second to leg pain, and the third to neurological symptoms. Numbers are number of patients per category (total n = 44) per timepoint, i.e., pre-surgery baseline, 3 and 12 months post-surgery. "Total agreement" cells are highlighted.

Physician rating (chart)

Baseline

| | +++ | ++– | +–+ | +–– | –++ | –+– | ––+ | ––– |
|---|---|---|---|---|---|---|---|---|
| +++ | **20** | 8 | 4 | 1 | 0 | 0 | 0 | 0 |
| ++– | 0 | **1** | 2 | 0 | 0 | 0 | 0 | 0 |
| +–+ | 0 | 1 | **0** | 0 | 0 | 0 | 0 | 0 |
| +–– | 3 | 0 | 1 | **0** | 0 | 0 | 0 | 0 |
| –++ | 2 | 1 | 0 | 0 | **0** | 0 | 0 | 0 |
| –+– | 0 | 0 | 0 | 0 | 0 | **0** | 0 | 0 |
| ––+ | 0 | 0 | 0 | 0 | 0 | 0 | **0** | 0 |
| ––– | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **0** |

3 months after surgery

Patient rating (questionnaire)

| | +++ | ++– | +–+ | +–– | –++ | –+– | ––+ | ––– |
|---|---|---|---|---|---|---|---|---|
| +++ | **2** | 1 | 2 | 3 | 3 | 0 | 1 | 1 |
| ++– | 0 | **0** | 1 | 0 | 0 | 1 | 0 | 0 |
| +–+ | 0 | 2 | **1** | 1 | 0 | 0 | 0 | 1 |
| +–– | 0 | 2 | 1 | **5** | 0 | 0 | 0 | 0 |
| –++ | 1 | 1 | 0 | 2 | **0** | 1 | 1 | 2 |
| –+– | 0 | 1 | 0 | 1 | 0 | **0** | 0 | 0 |
| ––+ | 1 | 0 | 0 | 0 | 0 | 0 | **0** | 0 |
| ––– | 1 | 0 | 0 | 2 | 0 | 0 | 0 | **2** |

12 months after surgery

| | +++ | ++– | +–+ | +–– | –++ | –+– | ––+ | ––– |
|---|---|---|---|---|---|---|---|---|
| +++ | **5** | 1 | 1 | 2 | 2 | 5 | 1 | 2 |
| ++– | 0 | **0** | 1 | 2 | 0 | 1 | 0 | 2 |
| +–+ | 0 | 1 | **0** | 2 | 0 | 0 | 0 | 1 |
| +–– | 0 | 0 | 0 | **3** | 0 | 0 | 1 | 0 |
| –++ | 1 | 1 | 0 | 0 | **0** | 1 | 0 | 2 |
| –+– | 0 | 0 | 0 | 0 | 0 | **0** | 0 | 0 |
| ––+ | 0 | 0 | 0 | 2 | 0 | 0 | **0** | 0 |
| ––– | 1 | 1 | 0 | 0 | 0 | 0 | 0 | **2** |

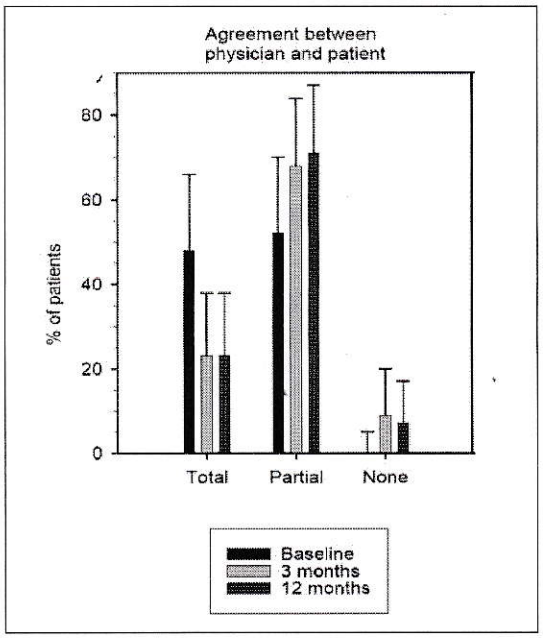


**Figure 1**

Agreement on symptoms as rated by physician (chart) and patient (questionnaire) before and 3 and 12 months after surgery (error bars are exact 95% confidence intervals).

leg pain were combined and determined again using the total agreement measure and Cohen's kappa based on two categories back/leg and neurological symptoms. Cohen's kappa = –0.12 at baseline (as many cases agreed with ++ rating), 0.22 at 3 months and 0.08 at 12 months, and the total agreement 61.4% at baseline, 47.7% at 3 months and 25.0% at 12 months. These values are larger than the corresponding ones based on three categories for 3 and 12 months after surgery.

The average age of patients with total, partial and no agreement at the twelve month post-surgery visit was 60, 64 and 68 years (n's = 11, 28 and 5; standard deviations = 13–15).

Symptoms before and after surgery were analysed separately for the three categories back-buttock pain, leg pain and neurological/disability. These results are summarised in figure 2. This shows that symptoms in all three categories were reduced significantly 3 and 12 months after surgery, especially leg pain and neurological symptoms. However, after surgery, physicians reported less leg pain and neurological signs than the patients did.

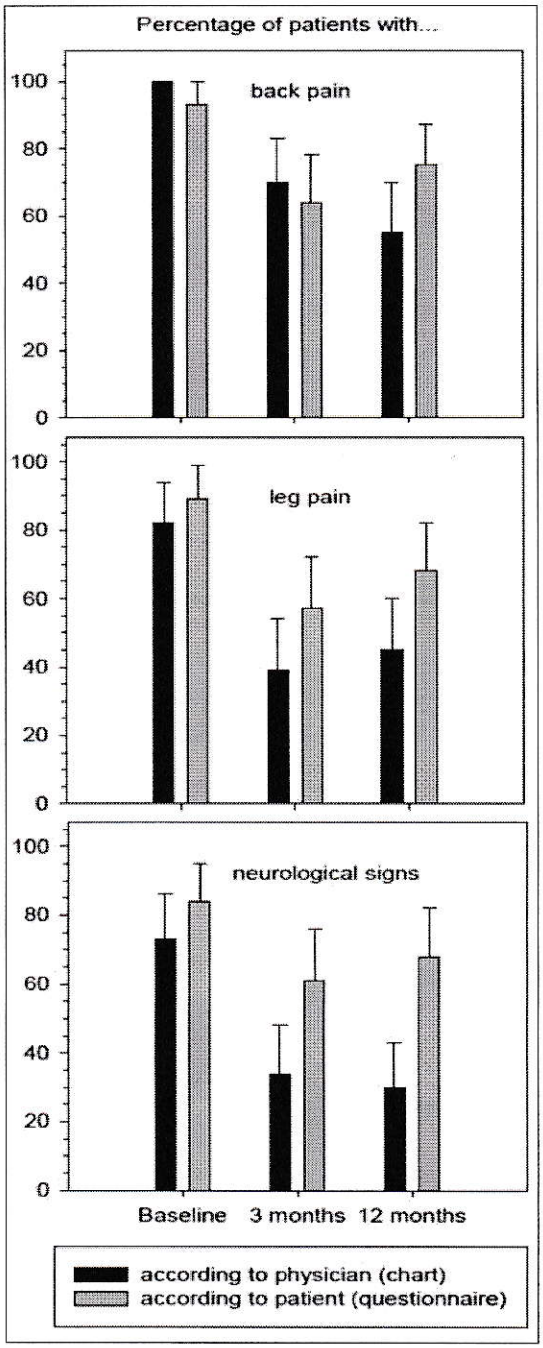


**Figure 2**

Agreement between doctor and patient by symptom (back/buttock pain, leg pain, neurological symptoms; error bars are Blyth-Still-Casella 95% confidence intervals).

## Discussion

Total agreement between patient and physician ratings of symptoms was substantially lower after surgery than before. This was mainly attributable to physicians reporting less leg pain and neurological signs than the patients did. There are several factors which could possibly contribute to this discrepancy between patient and physician appraisal.

Firstly, symptoms in the post-surgery consultations were less intense than before surgery. It is possible that patients rate milder symptoms more highly than the physician, while more intense symptoms are rated the same by both patient and physician. A similar effect has been reported after hip arthroplasty, whereby patient-physician rating differed more for moderate than severe pain [5].

Secondly, after a successful operation the surgeon will naturally expect a positive outcome, and may rate their results more highly than the patient; for example they may tend to ask only about

the improvement of symptoms, ignoring reported minor pain, and not asking about unchanged stronger pain. Other physician/patient comparative studies have also noted that patients report more pain than physicians, specifically after hip arthroplasty [3, 5, 10] and after shoulder arthroplasty [12].

Thirdly, it was found that younger patients have a higher rate of total or partial agreement compared to older patients. A similar age-related disagreement between physician and patient was reported by McGee et al. for hip arthroplasty [5]. It is possible that quality of communication is age-dependent, for example it could be that misinterpretation and misunderstanding are more frequent in older patients [5].

## Conclusion

The data presented in this study show that in clinical practice, the documentation in the patient's chart does not always accurately represent the patient's views.

If the patient's assessment in the questionnaire is accepted as correct (which it should be, given that the data are from a "gold standard" validated instrument), then it is the physician's routine assessment during outpatient clinics which needs to be improved. The fact that the discrepancy is larger for older patients suggested to us that more consultation time may be necessary with this group to elicit a truer picture of the patient's subjective symptoms. McGuirk and Bogduk have concluded in low back pain patients that 50 minutes are required for the first consultation [6]. This is a problem because our consultations are normally scheduled to take 15 minutes per patient including interview, examination and x-ray analysis. Based on the results of this study, it is concluded that it is necessary to increase the time allocated for consultations, and to more systematically structure the way the physician gathers information. It is hoped that this will help to ensure that the physician's patient chart more closely reflects the patient's self-rating, than was the case in this study.

*Correspondence:*
*Dr. Joerg Huber*
*Orthopedic Clinic*
*Kantonsspital*
*Tellstr. 1*
*CH-5001 Aarau*
*Switzerland*
*E-Mail: joerg.huber@ksa.ch*

## References

1 Daltroy LH, Cats-Baril WL, Katz JN, Fossel AH, Liang MH. The North American spine society lumbar spine outcome assessment Instrument: reliability and validity tests. Spine. 199;21(6):741–9.
2 Kopec JA. Measuring functional outcomes in persons with back pain: a review of back-specific questionnaires. Spine. 2000;25(24):3110–4.
3 Lieberman JR, Dorey F, Shekelle P, Schumacher L, Thomas BJ, Kilgus DJ, Finerman GA. Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty. J Bone Joint Surg Am. 1996;78(6):835–8.
4 Mannion AF, Junge A, Grob D, Dvorak J, Fairbank JC. Development of a German version of the Oswestry Disability Index. Part 2: sensitivity to change after spinal surgery. Eur Spine J. 2006;15(1):66–73. DOI: 10.1007/s00586-004-0816-z.
5 McGee MA, Howie DW, Ryan P, Moss JR, Holubowycz OT. Comparison of patient and doctor responses to a total hip arthroplasty clinical evaluation questionnaire. J Bone Joint Surg Am. 2002;84-A(10):1745–52.
6 McGuirk B, Bogduk N. Evidence-based care for low back pain in workers eligible for compensation. Occup Med (Lond). 2007;57(1):36–42. DOI: 10.1093/occmed/kql102.
7 Morlock RJ, Ward RE, Nerenz DR. Spine surgeons accurately estimate the probability of favorable one-year postoperative lumbar surgery outcomes. Disease Management. 2002;5(1):51–9. DOI: 10.1089/109350702317377564.
8 Peters A, Sabariego C, Wildner M, Sangha Dagger O. Outcome sensitivity of the North American Spine Society Instrument with special consideration for the neurogenic symptoms of chronic back pain. [Article in German] Z Orthop Ihre Grenzgeb. 2004;142(4):435–41. DOI: 10.1055/s-2004-820343.
9 Pose B, Sangha O, Peters A, Wildner M. Validation of the North American Spine Society Instrument for assessment of health status in patients with chronic backache. [Article in German] Z Orthop Ihre Grenzgeb. 1999;137(5):437–41.
10 Ragab AA. Validity of self-assessment outcome questionnaires: patient-physician discrepancy in outcome interpretation. Biomed Sci Instrum. 2003;39:579–84.
11 Schochat T, Rehberg W, von Kempis J, Stucki G, Jackel WH. The North American Spine Society Lumbar Spine Outcome Assessment Instrument: translation and psychometric analysis of the German version in rehabilitation patients with chronic back pain. [Article in German] Z Rheumatol. 2000;59(5):303–13. DOI: 10.1007/s003930070052.
12 Smith AM, Barnes SA, Sperling JW, Farrell CM, Cummings JD, Cofield RH. Patient and physician-assessed shoulder function after arthroplasty. J Bone Joint Surg Am. 2006;88(3):508–13. DOI:10.2106/JBJS.E.00132.
13 Pengel LHM, Refshauge KM, Maher CG. Responsiveness of pain, disability, and physical impairment outcomes in patients with low back pain. Spine;29(8):879–83.