Cold pain hypersensitivity predicts trajectories of pain and disability after low back surgery

a prospective cohort study

Müller, Monika; Bütikofer, Lukas; Andersen, Ole K.; Heini, Paul; Arendt-Nielsen, Lars; Jüni, Peter; Curatolo, Michele

Published in:
Pain

DOI (link to publication from Publisher):
10.1097/j.pain.0000000000002006

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

PAIN Publish Ahead of Print
DOI: 10.1097/j.pain.0000000000002006

Cold pain hypersensitivity predicts trajectories of pain and disability after low back surgery: a prospective cohort study

Monika Müller [1,2], Lukas Bütikofer [2,3], Ole K. Andersen [4], Paul Heini [5], Lars Arendt-Nielsen [4], Peter Jüni [6, 7], Michele Curatolo [8]

1. Translational Research Center, University Hospital of Psychiatry, University of Bern, Switzerland
2. CTU Bern, University of Bern, Switzerland
3. Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland
4. Center for Neuroplasticity and Pain (CNAP), Department of Health Science and Technology, School of Medicine, Aalborg University, Aalborg, Denmark
5. Department of Orthopedics, Private Clinic Sonnenhof, Bern, Switzerland
6. Applied Health Research Centre (AHRC) Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, Canada
7. Department of Medicine and Institute of Health Policy, Management and Evaluation, University of Toronto, Canada
8. Department of Anesthesiology and Pain Medicine, University of Washington, Seattle, USA

Correspondence:

Michele Curatolo, MD, PhD, University of Washington, Department of Anesthesiology & Pain Medicine, 1959 NE Pacific Street, Box 356540, Seattle, WA 98195-6540, USA

Telephone: +1-206 543 2568 Fax: +1-206 543 2958, E-Mail: curatolo@uw.edu

Number of text pages of the manuscript (including pages containing figures and tables): 32.

Number of figures: 3.

Number of tables: 3.

Abstract

Improving the ability to predict persistent pain after spine surgery would allow identification of patients at risk and guide treatment decisions. Quantitative sensory tests (QST) are measures of altered pain processes, but in our previous study preoperative QST did not predict pain and disability at single time-points. Trajectory analysis accounts for time-dependent patterns. We hypothesized that QST predict trajectories of pain and disability during one year after low back surgery.

We performed a trajectory analysis on the cohort of our previous study (n=141). Baseline QST included electrical, pressure, heat and cold stimulation of the low back and lower extremity, temporal summation and conditioned pain modulation. Pain intensity and Oswestry Disability Index were measured before, 2, 6 and 12 months after surgery. Bivariate trajectories for pain and disability were computed using group-based trajectory models. Multivariable regressions were used to identify QST as predictors of trajectory-groups, with sociodemographic, psychological and clinical characteristics as covariates.

Cold pain hypersensitivity at the leg, not being married and long pain duration independently predicted worse recovery (complete-to-incomplete, incomplete-to-no recovery). Cold pain hypersensitivity increased the odds for worse recovery by 3.8 (95% CI 1.8−8.0, p<0.001) and 3.0 (1.3-7.0, p= 0.012) in the univariable and multivariable analyses, respectively.

Trajectory analysis, but not analysis at single time-points, identified cold pain hypersensitivity as strong predictor of worse recovery, supporting altered pain processes as predisposing factor

for persisting pain and disability, and a broader use of trajectory analysis. Assessment of cold pain sensitivity may be a clinically applicable, prognostic test.

Keywords: Low back pain; failed back surgery syndrome; quantitative sensory tests; prediction

Introduction

Up to one third of patients complain of persistent low back pain and disability after spine surgery [9,11,22,28,37,79]. Most of these patients do not respond to conservative treatment or repeat surgery [11,34,42,48,56], which results in low quality of life and high unemployment rate [11,46,73]. Given the high prevalence of chronic low back pain [2,27,35,36,58], the high number of surgery for this condition [20,21], the high costs of surgery [18] and the burden of persistent pain and disability, it is crucial to improve our ability to identify patients at risk of poor recovery.

Alterations in pain processing resulting in pain amplification have been widely documented in animal studies and confirmed by human investigations in chronic low back pain using mechanistic quantitative sensory tests (QST) [7,30,53,57]. If altered pain processing of patients with chronic low back pain is a predisposing factor for persistent post-surgical pain, QST may contribute to predict the outcome of spine surgery.

Previous research on persistent impairment after spine surgery examined recovery at a single follow-up time point, and thus did not investigate time-depending patterns of recovery. In a previous analysis, we found none of 14 QST parameters to be predictive of pain and disability 12 months after spine surgery [51]. Trajectory analyses account for time-dependent patterns of disease development [39,52]. Studies adopting this method are emerging in pain research and have identified different trajectories for low back pain [6,12,19,23,25,39,40,72]. Because

trajectory analysis is able to reflect the dynamic nature of low back pain [5,39], it could identify predictors that are missed in prognostic studies using outcomes at single time-points.

Based on these premises, we performed a trajectory analysis on the cohort of our previous study [51], using follow-up time-points of 2, 6 and 12 months. The aim was to determine the ability of QST to predict the course of pain and disability during one year after spine surgery.

Methods

Study population and study setting

Patients undergoing up to three-level spine surgery for chronic low back pain associated with degenerative changes of the lumbar spine were recruited at three surgical tertiary care centers in Bern, Switzerland. Two assessors performed a clinical examination to confirm study eligibility. Patients with planned surgery for lumbosacral radiculopathy due to herniated discs, cancer or trauma were ineligible because clinical characteristics, surgical techniques and prognosis are different in these conditions, as compared to low back pain associated with degenerative changes [18].

Chronic low back pain was defined as lumbar back pain of ≥3 on a numerical rating scale (NRS, 0 ”no pain” and 10 “worst pain imaginable”) at most days during the week and with a minimum duration of three months, with or without radiation to the leg. Exclusion criteria were bilateral pain below the knees; rheumatologic inflammatory diseases; neurologic co-morbidities affecting the neurological function of the lower extremity to be tested; psychiatric co-morbidities other than unipolar depressive disorder; previous instrumented spine surgery (i.e. total disc replacement or spinal fusion with pedicle screws, cages or internal splints); planned surgery of more than three segments; and multiple somatic co-morbidities. We also excluded patients who could not be contacted by phone or mail before surgery.

Two assessors performed all study-related procedures according to a previously applied, standardized, prospective protocol [51,53] at the Department of Anesthesiology and Pain Medicine of the University Hospital of Bern.

The protocol was approved by the local research ethics committee (no. 176/11) and conducted in accordance with the Declaration of Helsinki [77]. All patients gave written informed consent.

Assessment of pain intensity and disability

Pain intensity was assessed using the NRS the Oswestry Disability Index (ODI) to assess disability [26] at baseline, 2, 6 and 12 months after surgery. Maximum pain during the last 7 days and average pain during the last 24 hours were recorded. The ODI describes current back-related disability with a score ranging from 0 “no disability” to 100 “maximum disability” [26]. An ODI of 10-20 indicates minimal, 21-40 moderate and >40 severe disability.

Quantitative Sensory Tests (QST)

QSTs were performed at the extremity contralateral to the most painful area of the lower back and at the most painful area at the lower back as measures of generalized and localized pain sensitivity, respectively. In case of bilateral back pain, the testing extremity was selected randomly according to a computer-generated list.

Patients were lying in a bed in a quiet room, with a leg rest placed under the knees to obtain a 30° semi-flexion. All patients underwent a training session to familiarize themselves with the stimulation procedure before data collection was initiated. Two measurements were performed, and the mean value was considered for data analysis, except for the cold pressor test and the assessment of conditioned pain modulation (CPM), for which only one measurement was performed. The sequence of testing modalities was randomly assigned according to a computer-generated list to avoid bias as a result of testing order [32].

Single and repeated electrical stimulation

Bipolar surface Ag/AgCl-electrodes were placed caudal to the lateral malleolus (innervation area of the sural nerve). Using a computer-controlled constant current stimulator (NCS System, Evidence 3102 evo, Neurosoft, Russia), the current intensity was increased from 1 mA in steps of 1 mA until the electrical stimulus was perceived as painful, which was considered as pain detection threshold.

A) Single stimulation: Single electrical stimulation consisted of a train-of-five 1-ms square-wave impulse of an overall duration of 25 ms. This train-of-five is perceived as a single stimulus.

B) Repeated stimulation (temporal summation): Temporal summation was elicited by repetition at a fixed stimulus intensity, which causes an increase in pain perception [3,60]. The train-of-five stimulus was repeated five times with a frequency of 2 Hz at a constant intensity. The intensity was increased as described above to assess the pain threshold, i.e., the stimulus intensity at which the repeated stimulation was perceived as painful.

Pressure stimulation

Pressure pain detection and tolerance thresholds were determined by an electronic pressure algometer with a 1 cm^2 surface probe (Somedic, Hörby, Sweden) [8]. The pressure tests were applied to the center of the pulp of the 2^{nd} toe and the site of most pain at the back. Pressure was increased from 0 at a rate of 30 kPa/s to a maximum of 1000 kPa. The pain detection threshold was defined as the point at which the pressure sensation turned into pain and pain tolerance threshold as the point at which the patient felt the pain as intolerable. The patients had to press a button when reaching these points and the algometer displayed the corresponding pressure intensity. Whenever a patient did not press the button below 1000 kPa, we considered this value as threshold.

Heat and cold stimulation

Pain detection threshold to heat and cold stimulation were assessed with a thermode of a 30 x 30 mm surface (TSA-ll; Medoc, Ramat Yishai, Israel) [53]. The tests were performed at the lateral aspect of the leg (midway between the knee and the lateral malleolus) and at the site of most pain at the back. The temperature of the thermode was changed at a rate of 0.5 ºC/sec from 30 ºC to a maximum of 50.5 ºC and to a minimum of 0.0 °C for heat and cold pain, respectively, until the stimulus was perceived as painful. The patients had to press a button when reaching these points. Threshold values were truncated in case patients did not report pain at the maximum of 50.5°C or the minimum of 0.0°C, respectively, and in this case these values were considered as thresholds.

Cold pressor test

The cold pressor test assesses the pain response to a tonic cold painful stimulus. The device consisted of a container separated into an outer and an inner part by a mesh screen, containing ice-saturated water (1.5± 1 °C monitored with a thermometer). The mesh screen prevented direct contact between the ice (placed in the outer part) and the hand of the subject (placed in the inner part). We asked the patients to immerse their hand for two minutes and withdraw the hand when they considered pain as intolerable. The hand withdrawal time was recorded for data analysis. Whenever a patient did not perceive the stimulus as intolerable below two minutes, this was considered as withdrawal time.

Conditioned pain modulation (CPM)

CPM was assessed using the cold pressor test as conditioning stimulus and pressure pain detection threshold at the 2nd toe as test stimulus [13,17,24,61,71]. Both stimulus modalities are described above. The pressure threshold was measured before the application of the conditioning stimulus (PPT1) and immediately after two minutes of hand immersion (PPT2). Patients who experienced intolerable pain before two minutes elapsed could briefly retract

their hand from the cold water and re-immersed it until two minutes were reached. The absolute CPM was assessed as the thresholds difference between PPT2 and PPT1, and a difference >0 was considered as a positive CPM response.

Baseline and surgery-related covariates

The evaluation of socio-demographic covariates included age, gender (female vs male), education (high school or university degree vs lower education), working status (regular work including houseworkers vs no regular work) and civil status (married vs unmarried). The Beck Depression Inventory version 2 (BDI-ll) was used to assess depression [50], the State-Trait-Anxiety-Inventory (STAI) for anxiety [44] and the Pain Catastrophizing Scale (PCS) [49,70] to determine the degree of catastrophizing.

The following clinical variables were recorded: Body-Mass-Index (BMI), smoking status (yes vs no), finger-ground distance (>10 cm vs ≤ 10 cm), positive Lasègue sign (yes vs no), previous non-instrumented back surgery (yes vs no), pain radiating to the leg (yes vs no), pain duration (years), as well as intake of non-opioid and opioid analgesics (yes vs no). Non-opioid analgesics were non-steroidal anti-inflammatory drugs, acetaminophen and metamizole.

The following surgery-related variables were recorded: type of surgery (instrumented vs non-instrumented surgery), number of segments to be operated (multi-segmental vs uni-segmental), intensity of acute post-surgical pain at the first day after surgery and at the last day before discharge using the maximum pain experienced on that day on the NRS. The average value of these two days was used for analysis.

Clinical management

Senior surgeons based the decision on the type of surgery and the number of segments to be operated upon clinical reasoning and radiologic findings [62,79]. They performed all surgeries under standard general anesthesia. Post-surgical treatment was standardized for all patients and included perioperative pain control using intravenous patient-controlled analgesia,

prescription of non-opioid analgesics for at least two weeks after surgery and stepwise rehabilitation. Prescription of opioids at discharge was not standard of practice in the participating institutions, except for preoperative opioids. Immediate rehabilitation consisted of stabilizing muscle exercises for trunk muscles in supine position according to a handout and encouragement of walking as much as tolerated. A rehabilitation training guided by the physical therapist begun two months after surgery.

Statistical analysis

The sample size has been calculated for the previously published study on the same cohort [51], based on an expected a frequency of failed back surgery of 30% [9,22,28]. A sample size of 155 patients can detect a dichotomized predictor that is approximately twice as frequent in patients with failed back surgery, if the frequency of the predictor was 25% or more. For continuous predictors, this sample size would detect a difference between patients with and without failed back surgery of 0.5 standard deviation (SD) units (power of 80% and a two-sided alpha of 0.05). Time and resource constraints led us to close the study 14 patients (9%) short of the planned 155 [51].

To model recovery after spine surgery, we computed bivariate trajectories for maximum pain during the last 7 days and disability using group-based trajectory models (GBTM) with censored normal distributions [52]. The time point of follow-up was used as a time variable and its effect was modelled with polynomials of different orders in each group. All possible bivariate models with up to three groups and third-order polynomials were fitted and the best model with at least five patients in each group was chosen according to the Akaike information criteria [1]. Based on the trajectories of the selected model, patients were classified into three ordered groups of complete, incomplete or no recovery. The average trajectories were plotted, including model-based 95% confidence intervals (95% CI) for each group.

The QST were analyzed as potential predictors of these three trajectory-groups using ordinal logistic regression models, with sociodemographic, psychological, pain-related and surgical characteristics at baseline as covariates. The above listed predictors were used as their possible association with poor recovery after surgery has been discussed in previous studies [10,47]. Since acute post-surgical pain is an important predictor for persistent post-surgical pain in general [31], but also an intermediate outcome lying on the causal pathway between baseline predictors and long-term pain and disability, this variable was included in a sensitivity analysis rather than in the main model.

Education, working conditions, marital status, type of low back pain and finger-ground distance were dichotomized according to pre-specified criteria to facilitate a clinically meaningful interpretation. Heat and cold pain detection thresholds as well as hand withdrawal time of the cold pressor test were truncated and could not be analyzed as continuous variables. Therefore, these variables were dichotomized post-hoc, using the maximally attainable stimulus as cut-off for heat and cold pain detection threshold (< 50.5°C and > 0.0° C) and the maximum time of hand immersion as cut-off for hand withdrawal time (< 120 sec).

Imputed missing baseline predictors were multiplied using chained equations with logistic regression for binary and predictive mean matching for continuous or ordinal variables, respectively, and generated 100 imputed datasets [63,65,69]. Within these datasets, we also accounted for the uncertainty in the group allocation from the GBTM. Instead of a fixed group allocation, the probability to belong to each group was used to simulate the allocation for each of the 100 datasets, so called pseudo-class draws [75].

Ordinal logistic regression models were computed assuming proportional odds for the simultaneous comparison of no to incomplete recovery and of incomplete to complete recovery, and reported effects as odds ratios (ORs) for a worse recovery with model-based 95% CI. This analysis yields a single set of results: the ORs refer to “worse recovery”, i.e. no

recovery instead of incomplete recovery or incomplete instead of complete recovery. The regression models were calculated for each dataset and averaged using Rubin's rules [64]. First, univariable analyses were performed and the predictors associated at $p<0.10$ were selected for a multivariable model. In case of high correlation between two variables, the one with the stronger association was chosen for the multivariable model. To ensure comparability of continuous and binary variables, the effect for all continuous variables was expressed per two standard deviations (2 SD) change [29]. For continuous socio-demographic, psychological and pain-related predictors, the effect was expressed per 2 SD increase, and for continuous QST as per 2 SD decrease. ORs above one suggest that dysfunction in pain processing (i.e. lower thresholds after pressure, electrical and heat stimulation, higher thresholds after cold stimulation, shorter hand withdrawal time of the cold pressor test and impaired CPM) is associated with poor recovery.

Then, three sets of GBTM sensitivity analyses were performed to assess the robustness of our main bivariate trajectory model. First, the influence of missing pain and disability values was assessed at different follow-up time points, using multiple imputation as described above. The number of patients with different assignment to trajectory-groups before and after multiple imputation was defined. Second, separate, univariate GBTM were fitted for both pain and disability and the resulting univariate trajectory-groups were compared to evaluate the appropriateness of the bivariate trajectory model as main model. The Krippendorff's alpha with bias-corrected bootstrapped 95% CI based on 1000 replications was used to quantify agreement between the univariate trajectory-groups defined either by pain or disability [41]. A Krippendorff's alpha of 0 indicates no agreement, 1 perfect agreement and -1 total disagreement. Third, bivariate GBTM for pain and disability was performed using average pain instead of maximum pain as secondary pain outcome. Bivariate trajectory-groups of the sensitivity analyses were compared with those of the main analysis using Krippendorff's alpha. Finally, two sets of sensitivity analyses were performed to assess the robustness of our

regression analyses. First, univariable generalized ordered logit models were used to relax the proportional odds assumption when simultaneously comparing no to incomplete recovery and incomplete to complete recovery [76]. Second, as described above, acute post-surgical pain was included as potential predictor in a multivariable sensitivity model.

All statistical analyses were done in Stata 15 (StataCorp. 2017, Stata Statistical Software Release 15, College Station, TX, StataCorp LLC). Plots were prepared in R 3.4.3 (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/).

Results

Study flow and completeness of data

A total of 958 patients undergoing surgery were screened for chronic low back pain between 2012 and 2015 and found 392 patients (41%) to be eligible (Figure 1). The three most important reasons for exclusion were previous back surgery (173 patients, 31%), multilevel degenerative changes with planned surgery of more than three segment (149 patients, 26%), and neurologic or psychiatric co-morbidity other than unipolar depressive disorder (94 patients, 17%). Of 392 eligible patients, 72 could not be reached (18%) and 132 refused study participation (46%). We included and analyzed 141 patients, with 135 (96%), 140 (99%) and 137 (97%) patients presenting at the 2, 6 and 12-month follow-up, respectively. No patient was completely lost to follow-up. Baseline assessment including quantitative sensory testing took place before surgery. The median time between baseline assessment and surgery was 5 days (IQR 2-9 days). Data completeness for sociodemographic, psychological and pain-related characteristics was high and ranged from 91% to 100%, with missing data due to incompletely filled assessment forms. Part of data on pressure, heat and cold pain detection thresholds, hand withdrawal time of the cold pressor test and CPM were missing due to

logistic reasons. Even for these data, completeness of variables was high and ranged from 82% to 100%.

Trajectories of pain and disability

The trajectories of pain and disability were concordant and defined three possible clinical courses of recovery. Figure 2 displays bivariate trajectories with 95% CI bands of pain intensity and disability. Table 1 shows pain intensity and disability at baseline, 2, 6 and 12 months after surgery in the whole cohort and stratified according to the three trajectory-groups.

Only 22 patients (16%) completely recovered after surgery. Most patients had incomplete recovery (N=67, 48%). Both groups with complete and incomplete recovery showed marked reduction in pain and disability within 2 months after surgery, and thereafter a continuing but slow decrease. Fifty-two patients (37%) did not recover, with about the same pain and disability scores over time as at baseline. Statistics for the best models and a detailed specification of the selected model are presented in E-Supplement 1 (available at http://links.lww.com/PAIN/B122). E-Supplement 2 (available at http://links.lww.com/PAIN/B122)displays bivariate trajectories of pain and disability of each patient with the average trajectory in red.

Characterization of the study population

The baseline and surgery-related characteristics as well as QST according to bivariate trajectory-groups are shown in Table 2. Eighty-one (57%) of all patients were female and the mean age was 61.1 (SD 13.7). The majority of the patients was married (N = 92, 65%). Mean depression and catastrophizing scores of the study population were low, i.e., 11.3 (SD 6.6) and 17.7 (SD 10.8), respectively.

Most patients complained of back pain radiating to the leg (N = 119, 84%), and 113 (80%) reported an average pain at baseline of more than four on the NRS. This was also reflected by

55 (39%) and 75 (53%) of study participants reporting ODI values reflecting moderate and severe disability at baseline. Sixty-three (45%) of all patients were taking non-opioid and 25 (18%) opioid analgesic at baseline.

Twenty-eight (20%) patients had a previous non-instrumented back surgery. Ninety-six (68 %) were operated at a single, 34 (24 %) at two and 11 (8 %) at three segments, respectively. In 49 (35 %) patients, decompression without additional instrumental stabilization was performed. We did not encounter any surgical complications. Mean acute postsurgical pain was 5.7 (SD 2.5).

Predictors of trajectories

Figure 3 shows the forest plot of QST and baseline characteristics as predictors of bivariate trajectories after surgery based on univariable ordinal logistic regressions assuming proportional odds. Three QST variables were associated with worse recovery in the univariable analyses. Cold pain hypersensitivity at the leg and at the back was predictive of worse recovery with an OR of 3.85 (95% CI 1.85-8.02, p<0.001) and 2.13 (95% CI 1.10-4.11, p=0.024), respectively. As mentioned in the Methods section, worse (or poor) recovery means complete-to-incomplete and incomplete-to-no recovery. Low pressure pain threshold at the back showed an association with the trajectory-groups with an OR of 2.32 (95% CI 1.20-4.48, p=0.013).

Not being married, high depression, high catastrophizing, positive Lasègue sign, long pain duration, high pain and high disability at baseline, intake of opioid and non-opioid analgesics were predictors of poor recovery in univariable analyses. Neither instrumented nor multi-segmental surgery was associated with an increased risk of poor recovery in univariable ordinal logistic regressions. Maximum pain after surgery was highly associated with poor recovery with an OR of 6.90 (95% CI 3.15 – 15.1, p < 0.001).

The results of the multivariable ordinal logistic regression are reported in table 3. The model included all predictors with $p < 0.10$ in univariable regressions, except average pain over the last 24 hours (highly correlated with maximum pain over the last 7 days), cold pain at the back (highly correlated with cold pain at the leg) and acute post-surgical pain (sensitivity analysis). Cold pain sensitivity at the leg independently predicted poor recovery with an OR of 2.98 (95% CI 1.27 - 7.00, $p = 0.012$). Among the covariates, not being married and long pain duration independently predicted poor recovery.

A previously published analysis of the same cohort using persistent pain at the single time-point of 12 months after surgery as main outcome yielded largely negative results: the point estimates of all QST were scattered around one and all 95% CI included one as measure of no association [51]. Cold pain detection threshold at the leg, which was highly significant in the present trajectory analysis, showed a statistical trend with an an OR of 1.88 (95% CI 0.89 to 3.99) and a p-value of 0.10. None of the secondary analyses revealed significant associations of any QST with persistent pain and disability.

Sensitivity analyses

The bivariate trajectory model based on maximum pain during the last 7 days and ODI remained robust in all three sensitivity analyses. Multiple imputation of missing pain and disability values at different follow-ups had no effect on the three recovery groups, with a maximum of 5 out of 141 patients (3.5%) having a different group assignment after imputation. Univariate trajectories of either pain or disability as displayed in E-Supplement 3 (available at http://links.lww.com/PAIN/B122) defined the same recovery pattern as the main bivariate trajectories of pain and disability. The univariate trajectory-groups showed high concordance with a Krippendorff's alpha of 0.62 (95% CI 0.51-0.73). E-Supplement 4 (available at http://links.lww.com/PAIN/B122) shows bivariate trajectories of average pain intensity of last 24 hours and disability. Again, agreement of the bivariate sensitivity

trajectory-groups and the bivariate main trajectory-groups was high, with a Krippendorff's alpha of 0.85 (95% CI 0.77 – 0.91). The forest plot in E-Supplement 5 (available at http://links.lww.com/PAIN/B122) represents the results of univariable generalized ordered logit models for bivariate trajectories of maximum pain of last 7 days and ODI. We did not find evidence for a violation of the proportional odds assumption. As shown in E-Supplement 6 (available at http://links.lww.com/PAIN/B122), when including acute post-surgical pain in the sensitivity multivariable ordinal logistic regression model, cold pain sensitivity at the leg, not being married and long pain duration still independently predicted poor recovery. The strongest predictor was acute post-surgical pain.

Discussion

Main findings

Cold pain sensitivity was a strong predictor of trajectories of persistent pain and disability one year after surgery. This finding has implications for development and validation of clinically applicable prognostic tests. The trajectory analysis identified predictors that with the previous study on the same cohort did not predict pain and disability at 12-month follow-up [51]. This finding has methodological implications.

Quantitative sensory tests

Quantitative sensory tests (QST) explore aspects of pain perception and nociceptive processes in humans. When applied to uninjured and healthy body sites, responses to QST reflect pain and nociceptive processes very likely occurring in the central nervous system [4,16]. Thus, the finding of pain hypersensitivity with cold stimulation at the leg suggests that alterations in central pain processes render patients vulnerable to poor long-term surgical outcome. In the univariable analyses, also cold pain hypersensitivity at the back and pressure pain threshold at the back were associated with poor recovery.

The results are supported by previous studies. In longitudinal studies on cervical trauma by Sterling et al, baseline cold pain threshold was the only QST variable that predicted persisting pain 6 and 12 months post-trauma [67,68], as well as trajectories of poor recovery [66]. In a cohort study on lateral epicondylalgia, baseline cold pain threshold was the only consistent predictor of pain, function and mechanical hyperalgesia at 12-month follow-up [15]. In a large cohort study using machine learning methodology, pain tolerance to hand immersion in cold water was strongly associated with lack of persistent pain 3 years after breast surgery, with a negative predictive value of 94.4% [45].

In our previous case-control study, several QST were able to discriminate patients with chronic low back from pain-free controls [53]. Cold pain thresholds were discriminative for all four body sites where the stimulation was applied, which included leg and back as in the present investigation. The areas under the receiver operating characteristic curves (ROC) for these four measurements were 0.62 to 0.76. The ROC for pressure pain threshold at the back, which was predictive in the univariate analysis of the present study, was 0.87 (95% CI 0.81–0.94). Early life stressors are associated with increased cold pain sensitivity, but not increased pressure pain sensitivity, at age 22 years, suggesting that cold modalities may be particularly sensitive in detecting the influence of socio-environmental and biologic factors on the development of pain and nociceptive hypersensitivity [74].

Little is known about the specific pain and nociceptive processes associated with different stimulus modalities. Therefore, there is no clear explanation for the predictive effect of cold stimulation as compared with other QST modalities. This question may be addressed by imaging studies elucidating the neural correlates of different stimulation modalities and their association with pain outcomes.

Covariates

Not being married, long pain duration and acute post-surgical pain independently predicted poor recovery. Acute post-surgical pain showed the strongest association with poor recovery with an OR of 5.88 (95% CI 2.34 - 14.8). These findings are in line with results of other studies that investigated persistent impairment after spine surgery [14,35,48].

Trajectory analysis

To our knowledge, the present study is the first one that determined trajectories of recovery after low back pain surgery. Previous cohorts investigated trajectories of low back pain in primary care patients [6,12,19,23,25,39,40,72]. Most of them observed three or four distinct trajectories that included clusters of complete recovery, incomplete recovery with either constant or fluctuating pain, and no recovery with persistently high pain. We found the similar recovery patterns.

The use of a trajectory analysis allowed the identification of predictors that had not been significantly associated with pain or disability at 12 months in the same cohort [51], in which no QST showed a statistically significant association with pain and disability at 12 months (main analysis). We have interpreted the statistical trend of cold pain sensitivity in the main analysis ($p=0.10$) and the statistical significance in one of the sensitivity analyses ($p=0.04$) as chance finding in view of 126 statistical tests performed [51].

The different findings of the previous and current analysis are likely explained by the different statistical approaches and outcomes. In the previous analysis, persistent pain at a single follow-up time-point was analyzed using logistic regression models. While this is a standard approach in prognostic studies, there is evidence that the prognosis of low back pain may not be adequately characterized by defining recovery at a single time-point and using a binary outcome measure [5,12,19,23,39]. Conversely, trajectory analyses account for patterns evolving over time, which enabled us to include pain and disability at all time-points and may

give a better picture of the course of recovery. While previously only two groups were used (recovery / non-recovery), introducing an intermediate group improved the resulting trajectory models substantially. Forcing the incomplete recoveries in either of the other two groups is likely to increase variance, reducing the power of the final predictor analysis. We also used a stochastic rather than a fixed group assignment, i.e. patients were allowed to belong to recovery groups with probabilities other than one. This procedure reduces the influence of patients with uncertain group assignments (as they contribute to more than one group), which may reduce variance and increase power.

Clinical course of pain and disability

An interesting question is whether pain and disability have a different course after spine surgery. We found a high overlap between pain and disability during follow-up, consistent with a previous cohort study in patients with acute low back pain, who experienced the same recovery pattern for pain and disability [19].

The most marked reduction in pain and disability was observed within two months after surgery, with a continuing but slow decrease thereafter (figure 2). However, 40% of the participants did not benefit from surgery with persistent and relatively high pain and disability.

Strengths and limitations

We consider the use of recovery trajectories with both pain and disability to define these trajectories as major strengths. Other strengths include the long follow-up period with three assessments, the near complete follow-up at all time-points, the comprehensive baseline predictors, and the excellent data quality with less than 10% missing data for all predictors except for CPM (18%). GBTM is a flexible statistical approach that takes advantage of the full longitudinal information on more than one clinical outcome [52].

The statistical approach to define trajectories is data-driven. This may limit generalizability, especially in the lack of external validation. We limited the maximum number of trajectory groups (three), the highest polynomial grades (three) and the minimum number of patients per group (five). This can be seen as arbitrary but ensures clinical interpretability.

An extensive, multimodal QST protocol was applied, using in total 14 tests to assess different dimensions of nociception and pain experience [54]. Five stimulation modalities were applied (electrical, pressure, heat dynamic, cold dynamic and cold tonic). Pain facilitation and inhibition were studied with a temporal summation and CPM model, respectively. Such an extensive protocol is another major strength. On the other hand, the evaluation of this large number of predictors increased the chance to detect false positive associations. Therefore, the reported p-values have to be interpreted accordingly. To base both univariable and multivariable regressions on the total study sample, multiple imputation was used to account for missing data [63,65,69]. The multivariable model examined eleven predictors with $p<0.10$, which is at the upper limit for the available number of patients [59]. However, precision of the estimates was still reasonable, and unexpected differences between the univariable and multivariable models were not observed.

Implications

The association of cold pain hypersensitivity with trajectories of pain and disability was clinically relevant. This finding justifies further studies to establish the prognostic value of cold pain testing. Noticeably, cold pain test can be delivered very quickly, making it suitable for large multi-site validation studies and of potential clinical use. The availability of a clinically applicable prognostic QST may improve our ability to reduce the failure rate of spine surgery with associated pain, disability and socioeconomic consequences. This could be accomplished by considering treatment alternatives for patients at risk. Pre- and peri-operative pain management could be optimized, aiming to reduce pain sensitivity by pharmacological

[38,55,78] and behavioral-health [14,43] interventions. Close follow-up could be planned to identify lack of recovery early and enroll patients for advanced treatments, such as multidisciplinary programs [33]. In addition, patients at risk of developing long term pain and disability could be selected for investigation aiming to develop and test preventive strategies.

A second implication is of methodological nature. The identification of predictors with the trajectory analysis, but not with the analysis on a single time-point on the same cohort, suggests that trajectory analyses should have broader use to improve our ability to identify significant predictors and detect time-dependent patterns of outcomes.

Finally, the most marked change in pain and disability, or lack thereof, occurred during the first two months after spine surgery. This finding suggests that preventive strategies aiming to shift the clinical course from a poor to a favorable recovery should be evaluated and implemented during the first two months.

Acknowledgements

We owe our special gratitude to Katrin Ziegler, MSc, Clinical Trials Unit, University of Bern, Switzerland, for her support in data-management; Fabienne Treichel, study nurse, for her support in patient recruitment and data collection; Prof. Andreas Raabe, University Clinic of Neurosurgery, Inselspital Bern, Switzerland and Dr. Ulrich Seidel, University Clinic of Orthopedics and Traumatology, Inselspital Bern, Switzerland, for their support in patient recruitment.

Funding: The study was funded by the Swiss National Science Foundation (No. 32003B_138361) and the Scientific Funds of the University Department of Anesthesiology and Pain Therapy of the University of Bern. Ole K. Andersen and Lars Arendt-Nielsen are supported by the Danish National Research Foundation (DNRF121).

Conflict of interests: No author has any conflicts of interest to declare.

Figure legends

Figure 1: Flow chart of study participant recruitment and follow-up.

[a] NRS: Numerical Rating Scale from 0 (no pain) to 10 (worst pain imaginable).

[b] Two patients with multiple sclerosis, two with dementia, one with postpolio-syndrome and one with epilepsy.

[c] Other: two withdrew consent, 6 had poly-morbidity.

Figure 2: Bivariate trajectories of maximum pain of the last 7 days and disability, defining three groups of recovery after spine surgery.

Data are presented as average effect for each group with 95% confidence bands. N=141. Group-based multi-trajectory model with polynomial grades 2, 3, 3, for both pain and disability.

£ Maximum pain during the last 7 days assessed by Numerical Rating Scale (0: no pain, 10: worst pain imaginable).

$ Disability assessed by Oswestry Disability Index (0: no disability to 100: maximum disability).

<u>Figure 3:</u> Forest plot of quantitative sensory tests (QST), socio-demographic, psychological and pain-related characteristics as predictors of bivariate trajectories after spine surgery.

Results are odds ratios with corresponding 95% confidence intervals (95% CI) and p-values of univariable ordinal logistic regressions*. N=141.

* Model assuming proportional odds when comparing “incomplete recovery” with “complete recovery” and “no recovery” with “incomplete recovery”, with effects of continuous predictors expressed per two standard deviation change (per 2 SD)

OR > 1.0 means altered quantitative sensory tests are associated with increased risk for poor recovery (i.e. lower thresholds after electrical, pressure and heat stimulation, higher thresholds after cold stimulation, shorter hand withdrawal time and impaired conditioned pain modulation).

References

[1] Akaike H. A New Look at the Statisical Model Identification. In: Pranzen E., Tanabe K., Kitagawa G. Springer Series in Statistics 1974:215-222.

[2] Andersson GB. Epidemiological features of chronic low-back pain. Lancet 1999;354:581-585.

[3] Arendt-Nielsen L, Brennum J, Sindrup S, Bak P. Electrophysiological and psychophysical quantification of central temporal summation of the human nociceptive system. Eur J Appl Physiol 1994;68:266-273.

[4] Arendt-Nielsen L, Morlion B, Perrot S, Dahan A, Dickenson A, Kress HG, Wells C, Bouhassira D, Mohr Drewes A. Assessment and manifestation of central sensitisation across different chronic pain conditions. Eur J Pain 2018;22:216-241.

[5] Axen I, Bergstrom G, Bodin L. Using few and scattered time points for analysis of a variable course of pain can be misleading: an example using weekly text message data. Spine J 2014;14:1454-1459.

[6] Axen I, Bodin L, Bergstrom G, Halasz L, Lange F, Lovgren PW, Rosenbaum A, Leboeuf-Yde C, Jensen I. Clustering patients on the basis of their individual course of low back pain over a six month period. BMC Musculoskelet Disord 2011;12:99.

[7] Biurrun Manresa JA, Neziri AY, Curatolo M, Arendt-Nielsen L, Andersen OK. Reflex receptive fields are enlarged in patients with musculoskeletal low back and neck pain. Pain 2013;154:1318-1324.

[8] Brennum J, Kjeldsen M, Jensen K, Jensen TS. Measurements of human pressure-pain thresholds on fingers and toes. Pain 1989;38:211-217.

[9] 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;28:1913-1921.

[10] Celestin J, Edwards RR, Jamison RN. Pretreatment psychosocial variables as predictors of outcomes following lumbar surgery and spinal cord stimulation: a systematic review and literature synthesis. Pain Med 2009;10:639-653.

[11] Chan CW, Peng P. Failed back surgery syndrome. Pain Med 2011;12:577-606.

[12] Chen Y, Campbell P, Strauss VY, Foster NE, Jordan KP, Dunn KM. Trajectories and predictors of the long-term course of low back pain: cohort study with 5-year follow-up. Pain 2018;159:252-260.

[13] Chitour D, Dickenson AH, Le Bars D. Pharmacological evidence for the involvement of serotonergic mechanisms in diffuse noxious inhibitory controls (DNIC). Brain Res 1982;236:329-337.

[14] Chou R, Huffman LH. Nonpharmacologic therapies for acute and chronic low back pain: a review of the evidence for an American Pain Society/American College of Physicians clinical practice guideline. Ann Intern Med 2007;147:492-504.

[15] Coombes BK, Bisset L, Vicenzino B. Cold Hyperalgesia Associated With Poorer Prognosis in Lateral Epicondylalgia: A 1-Year Prognostic Study of Physical and Psychological Factors. The Clinical Journal of Pain 2015;31:30-35.

[16] Curatolo M. Diagnosis of altered central pain processing. Spine 2011;36:S200-204.

[17] Danneels LA, Vanderstraeten GG, Cambier DC, Witvrouw EE, De Cuyper HJ. CT imaging of trunk muscles in chronic low back pain patients and healthy control subjects. Eur Spine J 2000;9:266-272.

[18] Deyo RA. Back surgery--who needs it? The New England journal of medicine 2007;356:2239-2243.

[19] Deyo RA, Bryan M, Comstock BA, Turner JA, Heagerty P, Friedly J, Avins AL, Nedeljkovic SS, Nerenz DR, Jarvik JG. Trajectories of symptoms and function in older adults with low back disorders. Spine (Phila Pa 1976) 2015;40:1352-1362.

[20] Deyo RA, Mirza SK, Martin BI, Kreuter W, Goodman DC, Jarvik JG. Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. JAMA 2010;303:1259-1265.

[21] Deyo RA, Nachemson A, Mirza SK. Spinal-fusion surgery - the case for restraint. N Engl J Med 2004;350:722-726.

[22] Dimar JR, 2nd, Glassman SD, Burkus JK, Pryor PW, Hardacker JW, Carreon LY. Two-year fusion and clinical outcomes in 224 patients treated with a single-level instrumented posterolateral fusion with iliac crest bone graft. Spine J 2009;9:880-885.

[23] Downie AS, Hancock MJ, Rzewuska M, Williams CM, Lin CW, Maher CG. Trajectories of acute low back pain: a latent class growth analysis. Pain 2016;157:225-234.

[24] Dubner R, Ren K. Endogenous mechanisms of sensory modulation. Pain 1999;Supplement 6:S45-S53.

[25] Dunn KM, Jordan K, Croft PR. Characterizing the course of low back pain: a latent class analysis. American journal of epidemiology 2006;163:754-761.

[26] Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine 2000;25:2940-2952.

[27] Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, Wallace AS, Castel LD, Kalsbeek WD, Carey TS. The rising prevalence of chronic low back pain. Archives of internal medicine 2009;169:251-258.

[28] 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;26:2521-2532; discussion 2532-2524.

[29] Gelman A. Scaling regression inputs by dividing by two standard deviations. Stat Med 2008;27:2865-2873.

[30] Giesecke T, Gracely RH, Grant MA, Nachemson A, Petzke F, Williams DA, Clauw DJ. Evidence of augmented central pain processing in idiopathic chronic low back pain. Arthritis Rheum 2004;50:613-623.

[31] Gilron I, Vandenkerkhof E, Katz J, Kehlet H, Carley M. Evaluating the Association Between Acute and Chronic Pain After Surgery: Impact of Pain Measurement Methods. The Clinical journal of pain 2017;33:588-594.

[32] Grone E, Crispin A, Fleckenstein J, Irnich D, Treede RD, Lang PM. Test order of quantitative sensory testing facilitates mechanical hyperalgesia in healthy volunteers. The journal of pain : official journal of the American Pain Society 2012;13:73-80.

[33] Guzman J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C. Multidisciplinary rehabilitation for chronic low back pain: systematic review. BMJ 2001;322:1511-1516.

[34] Hazard RG. Failed back surgery syndrome: surgical and nonsurgical approaches. Clin Orthop Relat Res 2006;443:228-232.

[35] Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, Woolf A, Vos T, Buchbinder R. A systematic review of the global prevalence of low back pain. Arthritis and rheumatism 2012;64:2028-2037.

[36] Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, Williams G, Smith E, Vos T, Barendregt J, Murray C, Burstein R, Buchbinder R. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Annals of the rheumatic diseases 2014;73:968-974.

[37] Ibrahim T, Tleyjeh IM, Gabbar O. Surgical versus non-surgical treatment of chronic low back pain: a meta-analysis of randomised trials. Int Orthop 2008;32:107-113.

[38] Khurana G, Jindal P, Sharma J, Bansal K. Post Operative Pain and Long Term Functional Outcome Following Administration of Gabapentin & Pregabalin in Patients Undergoing Spinal Surgery. Spine 2014;39:E363-368.

[39] Kongsted A, Kent P, Axen I, Downie AS, Dunn KM. What have we learned from ten years of trajectory research in low back pain? BMC Musculoskelet Disord 2016;17:220.

[40] Kongsted A, Kent P, Hestbaek L, Vach W. Patients with low back pain had distinct clinical course patterns that were typically neither complete recovery nor constant pain. A latent class analysis of longitudinal data. Spine J 2015;15:885-894.

[41] Krippendorff K. Content Analysis: An Introduction to Its Methodology. Beverly Hills CA, Sage 1980.

[42] Kumar K, Taylor RS, Jacques L, Eldabe S, Meglio M, Molet J, Thomson S, O'Callaghan J, Eisenberg E, Milbouw G, Buchser E, Fortini G, Richardson J, North RB. Spinal cord stimulation versus conventional medical management for neuropathic pain: A multicentre randomised controlled trial in patients with failed back surgery syndrome. Pain 2007;132:179-188.

[43] Lamb SE, Hansen Z, Lall R, Castelnuovo E, Withers EJ, Nichols V, Potter R, Underwood MR. Group cognitive behavioural treatment for low-back pain in primary care: a randomised controlled trial and cost-effectiveness analysis. Lancet 2010;375:916-923.

[44] Laux L, Glanzmann P, Schaffner P, Spielberger CD. State-Trait-Angstinventar (STAI). Göttingen: Hogrefe Verlag, 1981.

[45] Lötsch J, Ultsch A, Kalso E. Prediction of persistent post-surgery pain by preoperative cold pain sensitivity: biomarker development with machine-learning-derived analysis. Br J Anaesth 2017;119:821-829.

[46] Manca A, Eldabe S, Buchser E, Kumar K, Taylor RS. Relationship between health-related quality of life, pain, and functional disability in neuropathic pain patients with failed back surgery syndrome. Value Health 2010;13:95-102.

[47] Mannion AF, Elfering A. Predictors of surgical outcome and their assessment. Eur Spine J 2006;15 Suppl 1:S93-108.

[48] Martin BI, Mirza SK, Comstock BA, Gray DT, Kreuter W, Deyo RA. Reoperation rates following lumbar spine surgery and the influence of spinal fusion procedures. Spine (Phila Pa 1976) 2007;32:382-387.

[49] Meyer K, Sprott H, Mannion AF. Cross-cultural adaptation, reliability, and validity of the German version of the Pain Catastrophizing Scale. Journal of psychosomatic research 2008;64:469-478.

[50] Morley S, de CWA, Black S. A confirmatory factor analysis of the Beck Depression Inventory in chronic pain. Pain 2002;99:289.

[51] Müller M, Limacher A, Agten CA, Treichel F, Heini P, Seidel U, Andersen OK, Arendt-Nielsen L, Jüni P, Curatolo M. Can quantitative sensory tests predict failed back surgery?: A prospective cohort study. Eur J Anaesthesiol 2019;36:695-704.

[52] Nagin DS, Jones BL, Lima Passos VL, E. TR. Group-based multitrajectory modeling. Statistical Methods in Medical Research 2016;27:2015-2023.

[53] Neziri AY, Curatolo M, Limacher A, Nuesch E, Radanov B, Andersen OK, Arendt-Nielsen L, Juni P. Ranking of parameters of pain hypersensitivity according to their discriminative ability in chronic low back pain. Pain 2012;153:2083-2091.

[54] Neziri AY, Curatolo M, Nuesch E, Scaramozzino P, Andersen OK, Arendt-Nielsen L, Juni P. Factor analysis of responses to thermal, electrical, and mechanical painful stimuli supports the importance of multi-modal pain assessment. Pain 2011;152:1146-1155.

[55] Nielsen RV, Fomsgaard JS, Nikolajsen L, Dahl JB, Mathiesen O. Intraoperative S-ketamine for the reduction of opioid consumption and pain one year after spine surgery: A randomized clinical trial of opioid-dependent patients. Eur J Pain 2019;23:455-460.

[56] North RB, Campbell JN, James CS, Conover-Walker MK, Wang H, Piantadosi S, Rybock JD, Long DM. Failed back surgery syndrome: 5-year follow-up in 102 patients undergoing repeated operation. Neurosurgery 1991;28:685-690; discussion 690-681.

[57] O'Neill S, Manniche C, Graven-Nielsen T, Arendt-Nielsen L. Generalized deep-tissue hyperalgesia in patients with chronic low-back pain. Eur J Pain 2007;11:415-420.

[58] Palmer KT, Walsh K, Bendall H, Cooper C, Coggon D. Back pain in Britain: comparison of two prevalence surveys at an interval of 10 years. BMJ 2000;320:1577-1578.

[59] Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373-1379.

[60] Price DD. Characteristics of second pain and flexion reflexes indicative of prolonged central summation. Exp Neurol 1972;37:371-387.

[61] Pud D, Granovsky Y, Yarnitsky D. The methodology of experimentally induced diffuse noxious inhibitory control (DNIC)-like effect in humans. Pain 2009;144:16-19.

[62] Resnick DK, Watters WC, 3rd, Mummaneni PV, Dailey AT, Choudhri TF, Eck JC, Sharan A, Groff MW, Wang JC, Ghogawala Z, Dhall SS, Kaiser MG. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 10: lumbar fusion for stenosis without spondylolisthesis. J Neurosurg Spine 2014;21:62-66.

[63] Rubin DB. Inference and Missing Data. Biometrika 1976;63:581-592.

[64] Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons 2004.

[65] Spratt M, Carpenter J, Sterne JA, Carlin JB, Heron J, Henderson J, Tilling K. Strategies for multiple imputation in longitudinal studies. American journal of epidemiology 2010;172:478-487.

[66] Sterling M, Hendrikz J, Kenardy J. Similar factors predict disability and posttraumatic stress disorder trajectories after whiplash injury. Pain 2011;152:1272-1278.

[67] Sterling M, Jull G, Kenardy J. Physical and psychological factors maintain long-term predictive capacity post-whiplash injury. Pain 2006;122:102-108.

[68] Sterling M, Jull G, Vicenzino B, Kenardy J, Darnell R. Physical and psychological factors predict outcome following whiplash injury. Pain 2005;114:141-148.

[69] Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393.

[70] Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychological Assessment 1995;74:524-532.

[71] Suzuki R, Rygh LJ, Dickenson AH. Bad news from the brain: descending 5-HT pathways that control spinal pain processing. Trends Pharmacol Sci 2004;25:613-617.

[72] Tamcan O, Mannion AF, Eisenring C, Horisberger B, Elfering A, Muller U. The course of chronic and recurrent low back pain in the general population. Pain 2010;150:451-457.

[73] Thomson S, Jacques L. Demographic characteristics of patients with severe neuropathic pain secondary to failed back surgery syndrome. Pain practice : the official journal of World Institute of Pain 2009;9:206-215.

[74] Waller R, Smith AJ, O'Sullivan PB, Slater H, Sterling M, Straker LM. The association of early life stressors with pain sensitivity and pain experience at 22 years. Pain 2020;161:220-229.

[75] Wang C-P, Hendricks Brown C, Bandeen-Roche K. Residual Diagnostics for Growth Mixture Models. Journal of the American Statistical Association 2005;100:1054-1076.

[76] Williams R. Generalized Ordered Logit/ Partial Proportional Odds Models for Ordinal Dependent Variables. The Stata Journal 2006;6:58-82.

[77] Worlds. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. . Seoul: Wordls Medical Association 2008.

[78] Yu L, Ran B, Li M, Shi Z. Gabapentin and Pregabalin in the Management of Postoperative Pain After Lumbar Spinal Surgery: A Systematic Review and Meta-analysis. Spine 2013;38:1947-1952.

[79] Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus non-surgical treatment for lumbar spinal stenosis. Cochrane Database Syst Rev 2016:CD010264.

Table 1: **Pain and disability during the study period. Results are mean (SD). N = 141.**

	All patients (n = 141)		Complete recovery (group 1, n = 22, 16%)		Incomplete recovery (group 2, n = 67, 48%)		No recovery (group 3, n = 52, 37%)	
Pain intensity (NRS) [£]								
	N*		N*		N*		N*	
Baseline	140	7.79 (1.37)	22	7.73 (1.03)	66	7.48 (1.50)	52	8.19 (1.22)
2 months	135	3.97 (2.67)	21	0.76 (1.14)	66	3.38 (2.14)	48	6.19 (1.82)
6 months	140	3.90 (3.01)	22	0.00 (0.00)	67	2.99 (1.99)	51	6.78 (1.91)
12 months	137	3.62 (3.11)	22	0.09 (0.29)	67	2.57 (2.15)	48	6.71 (2.06)
Disability (ODI) [$]								
Baseline	140	40.4 (12.8)	22	36.5 (15.5)	66	37.6 (11.5)	52	45.5 (11.8)
2 months	133	22.2 (15.3)	21	8.71 (9.9)	65	18.7 (12.2)	47	33.1 (14.1)
6 months	140	20.4 (17.9)	22	1.59 (4.01)	67	13.4 (8.92)	51	37.5 (16.1)
12 months	137	18.6 (16.1)	22	1.86 (3.47)	67	12.4 (8.24)	48	35.0 (13.8)

N* No. of patients with complete data of the corresponding variable.

[£] Maximum pain during the last 7 days. NRS: Numerical Rating Scale (0: no pain, 10: worst pain imaginable).

[$] Oswestry Disability Index (0: no disability to 100: maximum disability).

Table 2: Baseline characteristics, Quantitative Sensory Tests and surgery related characteristics per group defined by the bivariate trajectories based on maximum pain and disability after surgery. Results are number of patients (%), mean (SD) or median [lq, uq]. N = 141.

	Complete recovery (group 1, n = 22, 16%)		Incomplete recovery (group 2, n = 67, 48%)		No recovery (group 3, n = 52, 37%)	
Sociodemographic characteristics	N*		N*		N*	
Age (years)	22	61.6 (11.5)	67	61.4 (13.8)	52	60.6 (14.7)
Female	22	12 (55%)	67	39 (58%)	52	30 (58%)
Higher education	22	4 (18%)	67	18 (27%)	52	12 (23%)
Regular work #	22	10 (45%)	67	26 (39%)	52	19 (37%)
Married	22	17 (77%)	67	49 (73%)	52	26 (50%)
Psychological characteristics						
Depression (BDI-II)	22	8.23 (5.94)	66	10.7 (6.21)	52	13.3 (6.83)
Anxiety (STAI Trait)	21	53.1 (7.91)	65	53.7 (8.62)	52	55.5 (6.92)
Catastrophizing (PCS)	20	12.8 (11.9)	60	16.5 (9.22)	48	21.4 (11.2)
Pain-related characteristics						
Body mass index (kg/m2)	22	29.0 (4.60)	67	27.9 (4.82)	52	28.5 (4.08)
Smoking	22	4 (18%)	66	19 (28%)	51	16 (31%)
Large finger ground distance (>10 cm)	21	11 (50%)	65	33 (49%)	51	34 (65%)
Lasègue positive	21	8 (36%)	64	25 (37%)	52	33 (63%)
Previous low back surgery	22	4 (18%)	67	10 (15%)	52	14 (27%)
Low back pain with irradiation to leg	22	19 (86%)	65	56 (84%)	52	44 (85%)
Pain duration (years)	22	1.75 [0.83, 2.00]	63	2.00 [0.75, 5.00]	49	4.00 [1.50, 15.0]
Maximum pain over the last 7 days at baseline (NRS)	22	7.73 (1.03)	66	7.48 (1.50)	52	8.19 (1.22)
Average pain over the last 24 hours at baseline (NRS)	22	5.73 (1.45)	67	5.48 (1.49)	52	6.13 (1.55)
Disability (ODI)	22	36.5 (15.5)	66	37.6 (11.5)	52	45.5 (11.8)
Intake of non-opioid analgesics	22	7 (32%)	67	26 (39%)	52	30 (58%)
Intake of opioid analgesics	22	3 (14%)	66	5 (7.5%)	51	17 (33%)
Quantitative Sensory Tests						
Electrical pain detection threshold single stimulation (mA)	22	9.8 (4.29)	67	9.35 (4.24)	52	9.48 (5.65)

Electrical pain detection threshold repeated stimulation (mA)	22	6.80 (2.60)	67	6.37 (2.78)	52	6.54 (2.88)
Pressure pain detection threshold at 2nd toe (kPa)	22	272 (91.6)	67	269 (114)	52	264 (109)
Pressure pain detection threshold at back (kPa)	22	370 (198)	66	366 (179)	52	275 (146)
Heat pain at leg (<50.5°C)	22	15 (68%)	65	42 (63%)	50	41 (79%)
Heat pain at back (<50.5°C)	21	19 (86%)	65	54 (81%)	50	48 (92%)
Cold pain at leg (>0.0°C)	22	5 (23%)	65	14 (21%)	49	28 (54%)
Cold pain at back (>0.0°C)	22	9 (41%)	66	26 (39%)	49	32 (62%)
Cold pressor test hand withdrawal (<120 sec)	21	15 (68%)	65	57 (85%)	50	43 (83%)
Impaired conditioned pain modulation	18	2 (9.1%)	57	11 (16%)	40	7 (13%)
Surgery-related characteristics						
Instrumented surgery	22	11 (50%)	66	44 (66%)	52	36 (69%)
Multi-segmental surgery	22	10 (45%)	67	18 (27%)	52	17 (33%)
Acute post-surgical pain (NRS)	21	3.48 (2.24)	60	5.57 (2.40)	47	6.98 (1.81)

N* No. of patients with complete data of the corresponding variable

Includes houseworkers

BDI-ll: Beck Depression Inventory Version 2 (0: no depression to 63: maximum depression)

STAI: State Trait Anxiety Index

PCS: Pain Catastrophizing Scale (0: no catastrophizing to 52: maximum catastrophizing)

NRS: Numerical Rating Scale (0: no pain,10: maximum pain)

ODI: Oswestry Disability Index (0: no disability, 100: maximum disability)

Table 3: Predictors of bivariate trajectories based on maximum pain of the last 7 days and disability after surgery. Results are odds ratios (OR) with corresponding 95% confidence intervals (95 % CI), and p-values of a multivariable ordinal logistic regression*. N=141.

	OR (95% CI)	P-value
Cold pain at leg (>0.0°C)	2.98 (1.27 - 7.00)	0.012
Pressure pain detection threshold at back (per 2 SD decrease)	1.52 (0.72 - 3.24)	0.27
Married	0.36 (0.15 - 0.84)	0.017
Depression (per 2 SD increase)	1.18 (0.45 - 3.10)	0.73
Catastrophizing (per 2 SD increase)	1.92 (0.72 - 5.12)	0.19
Lasègue positive	1.13 (0.50 - 2.57)	0.77
Pain duration (per 2 SD increase)	3.56 (1.37 - 9.26)	0.009
Maximum pain over the last 7 days at baseline (per 2 SD increase)	0.73 (0.32 - 1.65)	0.45
Disability at baseline (per 2 SD increase)	1.67 (0.67 - 4.17)	0.27
Intake of non-opioid analgesics	2.15 (0.94 - 4.93)	0.07
Intake of opioid analgesics	1.96 (0.64 - 5.98)	0.24

* The model assumes proportional odds when comparing “incomplete recovery” with “complete recovery” and “no recovery” with “incomplete recovery” with effect of continuous predictors expressed per two standard deviation change (per 2 SD). The model includes all predictors with $p < 0.10$ in univariable regressions except average pain over the last 24 hours (highly correlated with maximum pain over the last 7 days); cold pain detection threshold at the most painful site of the back (highly correlated with cold pain detection threshold at the leg) and acute post-surgical pain (sensitivity analysis).
OR > 1.0 suggests association with poor recovery