Predictors of pain and disability outcomes following spinal surgery for chronic low back and radicular pain: A systematic review

Monika Halicka, PhD[a]*; Rui Duarte, PhD[b]; Sharon Catherall, MPh[c]; Michelle Maden, PhD[b]; Michaela Coetsee[a]; Martin Wilby, FRCS, PhD[d]; Christopher Brown, PhD[a]

[a]Department of Psychological Sciences, University of Liverpool, Liverpool, United Kingdom

[b]Liverpool Reviews & Implementation Group (LRiG), University of Liverpool, Liverpool, United Kingdom

[c]Public Health Policy and Systems / LRiG, University of Liverpool, Liverpool, United Kingdom

[d]Department of Neurosurgery, The Walton Centre NHS Foundation Trust, Liverpool, United Kingdom

***Corresponding author**

Monika Halicka

Email: mon.halicka@gmail.com

Phone: +44 (0)1517941398

Address: Department of Psychological Sciences, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool, L69 7ZA, United Kingdom

Conflicts of Interest and Source of Funding

This research was funded via the Translational Research Access Programme (TRAP), Faculty of Health & Life Sciences, University of Liverpool, UK. The funder approved the objective of this review but had no role in data collection and synthesis, decision to publish, or preparation of the manuscript. The authors have no conflicts of interest to declare.

Abstract

Objectives: Success rates of spinal surgeries to treat chronic back pain are highly variable and useable prognostic indicators are lacking. We aimed to identify and evaluate preoperative predictors of pain and disability after spinal surgery for chronic low back/leg pain (CLBP). **Methods:** Electronic database (01/1984-03/2021) and reference searches returned 2622 unique citations. Eligible studies included adults with CLBP lasting ≥3 months undergoing first elective lumbar spine surgery, and outcomes defined as change in pain (primary)/disability (secondary) after ≥3 months. We included 21 reports (6899 participants), 7 judged to have low and 14 high risk of bias. We performed narrative synthesis and determined the quality of evidence (QoE). **Results:** Better pain outcomes were associated with younger age, higher education, and no spinal stenosis (low QoE); lower preoperative pain, less comorbidities, lower pain catastrophizing, anxiety and depression (very low QoE); but not with symptom duration (moderate QoE), other sociodemographic factors (low QoE), disability, or sensory testing (very low QoE). More favorable disability outcomes were associated with preoperative sensory loss (moderate QoE); lower job-related resignation and neuroticism (very low QoE); but not with socioeconomic factors, comorbidities (low QoE), demographics, pain, or pain-related psychological factors (very low QoE). **Discussion:** In conclusion, absence of spinal stenosis potentially predicts greater pain relief and preoperative sensory loss likely predicts reduction in disability. Overall, QoE for most identified associations was low/very low.

Keywords

Chronic low back pain; spinal surgery; predictors; pain; disability

1. Introduction

Approximately 40% of the worldwide population will experience low back pain (LBP) in their lifetime [1]. While most acute episodes resolve within several weeks [2], over 60% of people with LBP are estimated to have persistent or recurring pain a year later [3]. Chronic LBP is the single greatest cause of years lived with disability worldwide [4]. Its rapidly rising prevalence is expected to increase further given an aging population, increase in obesity, and reduction in activity, which are significant risk factors for LBP [5]. Accordingly, the rates of surgeries to treat LBP secondary to spinal pathologies have approximately doubled in the US and UK over the previous decade [6,7]. While spinal surgery costs the UK National Health Service approximately £500 million annually [8], its success rates are highly variable. Only about 60% of patients undergoing index lumbar spine surgery achieve minimal clinically important reductions in pain intensity [9–11].

Reliable predictive factors have the potential to inform clinical decision making to help maximize patient benefit and cost-effectiveness, yet there are no clear guidelines on useful predictors. Common surgical indications include symptom severity, non-response to conservative treatment, and imaging evidence of underlying pathology [12,13]. However, regarding prognosis, the UK National Institute for Health and Care Excellence suggests not using factors such as BMI, smoking status, or psychological distress to select patients for spinal surgery due to insufficient high-quality evidence [12]. Therefore, a comprehensive synthesis and evaluation of evidence regarding predictors of spinal surgery outcomes for chronic LBP is warranted. Knowledge of pre-identified reliable prognostic factors could inform clinical decision making regarding the best course of treatment, and also guide individualized preoperative interventions targeting modifiable risk factors to optimize patient outcomes. For instance, fusion surgery for back pain has better outcomes if patients have successfully completed a pain management course with cognitive-behavioral therapy [14].

Previous systematic reviews addressing similar questions were restricted to specific pathologies such as disc herniation [15–19] or surgical interventions such as spinal fusion [20–22]. However, the prognostic value of sociodemographic, health-related, and psychological patient characteristics for reduction in pain and disability may be independent of medical diagnosis and type of surgery, and considering broader LBP population could potentially mitigate the issues of insufficient amount or quality of evidence faced by previous reviews. Notably, the potential impact of LBP duration appears overlooked, as except for a review from 2011 looking at predictors of differential response to fusion versus conservative treatment [20,21], none of the relevant systematic reviews in the field specifically considered individuals with chronic symptoms that may be more resistant to treatment. Indeed, there is little change in pain and disability over the course of LBP if the symptoms do not resolve within several weeks since their onset [2,23] and individuals who have been living with symptoms for longer show poorer response to LBP treatments [24,25]. Given these gaps in evidence, the current review aimed to identify and evaluate preoperative predictors of pain and disability outcomes after spinal surgery for the treatment of chronic LBP and/or radicular pain.

2. Methods

This systematic review was conducted and reported in accordance with the general principles outlined in the Centre for Reviews and Dissemination (CRD) guidance for conducting reviews in health care [26] and the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) [27]. A review protocol has been prospectively registered at PROSPERO (ref. CRD42020180845) prior to formal screening of search results against eligibility criteria.

2.1. Search strategy

The search strategy was developed in collaboration with an information specialist (MM). For full electronic search strategy, including notes on any limits and search filters, see **Text S1, Supplemental Digital Content 1**. Electronic database searches (MEDLINE, EMBASE, PsycINFO, CINAHL, Cochrane Central Register of Controlled Trials [CENTRAL]) were performed on 8 April 2020

and updated on 29 March 2021. Search results were exported to EndNote library and de-duplicated. We also manually searched the reference lists of included studies and relevant systematic reviews [15–22,28–30] to identify any additional primary studies [31].

2.2. Eligibility criteria

Table 1 summarizes the eligibility criteria in a modified PICOTS format for reviews of prognostic studies (Population, Index and Comparator prognostic factors, Outcomes, Timing, Setting) [32] with additional specification of eligible study designs and publication formats. For detailed justification for choosing specific criteria and decision rules in case of uncertainty, see **Text S2, Supplemental Digital Content 2**.

In line with our aim to identify relevant predictors of change in pain and/or disability after spinal surgery, we applied broad and comprehensive inclusion criteria, thus as *index prognostic factors*, we considered any baseline factors, assessed prior to surgery, investigated for their potential to predict these outcomes. There were no restrictions applied to *comparator prognostic factors*, defined as 'adjusted for' factors used to investigate the independent prognostic value of a particular index prognostic factor over and above other (comparator) factors, as we considered both unadjusted and adjusted prognostic effects, where available.

Our primary outcome was change in pain intensity measured as (a) proportion of patients achieving minimal clinically important difference (MCID, as defined by study authors) in back and/or leg pain intensity, or (b) the magnitude of reduction in back and/or pain intensity from baseline to the last available follow-up as a continuous score. We included 0-10 Numerical Rating Scale (NRS) and 0-100 Visual Analog Scale (VAS; scores can be transformed into a 0-10 scale) as recommended pain measures in LBP research [33,34]. Our secondary outcome included change in disability measured as (a) proportion of patients achieving MCID on Oswestry Disability Index (ODI [35]), Roland-Morris Disability Questionnaire (RMDQ [36]), or Core Outcome Measures Index (COMI [37]), (b) the magnitude of reduction in disability on these measures from baseline to the last available follow-up

as a continuous score, or (c) return to work. ODI and RMDQ were recommended as the core measures of physical functioning/disability outcomes in back pain research [33,34,38], and COMI has been adapted as multidimensional outcome measure by the European Spine Society [39]. In addition to these condition-specific outcome measures, we also considered non-specific functional measures such as Short Form Health Survey for narrative synthesis. Return to work was also included as an objective measure of functional improvement, however, this outcome will be reported in a separate manuscript. Eligible studies reported at least one of the above-mentioned outcomes with ≥3 months follow-up. Throughout the article, we refer to positive pain and disability outcomes, that is, achieving MCID or greater reduction in pain or disability, consistent with a success of or greater benefit from surgery.

2.3. Study selection

To limit any potential selection bias, two reviewers (MH and RD) independently screened titles and abstracts, and then full texts, against the eligibility criteria for inclusion in the systematic review. Any disagreements were resolved by discussion and consensus, and an opinion from a third reviewer (MW) was sought where necessary. Abstracts with uncertain eligibility were included in the full text screening. Custom screening and selection tables in MS Excel, piloted on five randomly selected full-text articles, were used to record the selection process and reasons for exclusion. If eligibility could not be determined with certainty based on the information provided in the full text, supplementary materials, or related publications, additional details were requested from the study's corresponding authors (n = 10), who were re-contacted after a week if no response was received. The selection process is outlined in a PRISMA flowchart [27] (**Figure 1**).

2.4. Data extraction

Data extraction followed the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prognostic Factor Studies (CHARMS-PF) suggesting the key items to be extracted from primary studies of prognostic factors [32]. Two independent reviewers (MH and SC) piloted the CHARMS-PF-based tool on 2 randomly selected included studies. Data was recorded in a data

extraction form in MS Excel, a template of which is provided in **Table S3, Supplemental Digital Content 3**. Each reviewer extracted the data from half of the included articles and checked the data extracted by the other reviewer for accuracy. Any disagreements between the reviewers' judgements were successfully resolved by discussion and consensus.

Where possible, we extracted the adjusted effects of prognostic factors from multivariate models, however, to retain as much of the available data as possible, we also separately extracted unadjusted prognostic effects from univariate models. We aimed to obtain common effect estimates for each type of outcome, that is, odds ratio (OR) for binary outcomes or standardized mean difference for continuous outcomes and confidence intervals or standard errors of these estimates, or correlation coefficients for continuous outcomes where unadjusted associations were reported. To avoid potential selection bias, if the desired effect estimates were not reported, we converted or calculated the desired ones based on available data (e.g. 2x2 tables) using effect size calculators [40,41].

2.5. Risk of bias assessment

The risk of bias (RoB) was assessed at study level, using Quality in Prognosis Studies (QUIPS) tool [42–44]. Each of 6 QUIPS domains (study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding, statistical analysis and reporting) was rated as being at high, moderate, or low RoB. Domain ratings were guided by prompting items based on criteria suggested by Grooten et al. [42], which we modified and elaborated for the purpose of the current review question. All available reports based on the same study were appraised separately where applicable. Following the calibration of the QUIPS form between 2 independent reviewers (MH and SC), additional criteria were specified that included degrading RoB in the study attrition domain for studies that retrospectively recruited only patients who had complete follow-up data, as complete cases may be systematically different from eligible study sample. We specified the key characteristics of interest (age, sex, socioeconomic status, duration of symptoms, location of pain, underlying pathology, type of surgery) in study participation and attrition domains as relevant

to the current review question. The prompting item regarding the source of target population did not contribute to study participation domain RoB ratings as this is not commonly reported in the field under review. No specific set of required confounders was defined *a priori* as there is no established agreement on which factors should be included, but RoB ratings in the study confounding domain were downgraded in absence of any adjusted analyses. Finally, inadequate sample size or lack of power calculation were considered as potential sources of bias in the statistical analysis and reporting domain. The complete QUIPS template is available in **Table S4, Supplemental Digital Content 4**. Each reviewer independently rated the RoB for half of the included reports, and each checked the ratings of the other reviewer for agreement. Any disagreements were resolved by consensus. The overall RoB for each report was rated as 'low' if all six domains of QUIPS were judged to be at low-moderate RoB, or 'high' if one or more domains were judged to be at high RoB [45]. Results of this assessment were considered in the narrative synthesis and grading the level of evidence.

2.6. Data synthesis

Given sufficient and appropriate data for quantitative synthesis, we planned to perform meta-analyses of the effects of predictive factors on the primary and secondary outcomes. However, quantitative synthesis was not possible because many prognostic factors were only assessed in single studies and the remaining studies were too heterogeneous in terms of analysis types and outcome and predictor definitions. In particular, it was not feasible to combine effect estimates from studies using different analysis methods or reporting insufficient information to allow transformations, studies using continuous and dichotomous outcomes or predictors, different cut-offs for dichotomous outcomes, or different categorizations of the same predictors.

Therefore, we presented a tabular summary of adjusted and unadjusted associations between index prognostic factors and each outcome, accompanied by a narrative synthesis of the results. We summarized the number of studies that investigated relationships between each predictor and outcome, discussed the direction and strength of any associations and the consistency of evidence

across studies, and evaluated the findings considering the results of RoB assessment at the study and outcome level.

2.7. Grading of evidence

Two reviewers (MH and MC) simultaneously and collaboratively evaluated the strength of evidence for pain and disability outcomes using the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) tool [46] adapted for reviews of prognostic studies [47]. The overall quality of evidence for association of each category of predictors with each outcome was rated as high, moderate, low, or very low, reflecting the level of confidence that the true effect lies close to the estimate of the effect.

Phase of investigation determined the starting quality of evidence. This was considered high for phase-3 studies providing evidence of the mechanisms of action of the prognostic factor on the outcome, and for hypothesis-driven phase-2 studies testing independent associations of a prognostic factor with outcome. The starting quality was considered moderate for phase-1 studies exploring potential associations between prognostic factors and the outcomes, thus generating hypotheses about identified relationships [47]. Studies which investigated hypothesized associations but only in unadjusted analyses or for a class of multiple predictors, were classified as phase-1 studies. From this initial grade, the quality of evidence was downgraded for (a) serious study limitations when most evidence was from high RoB studies or from unadjusted analyses; (b) clinically meaningful inconsistency in results across studies (e.g. variable direction of or presence of a significant association) that could not be explained by differences in population characteristics, duration of follow-up, definitions of the predictor or outcome; (c) indirectness where the study sample, predictor, or outcome did not accurately reflect the review question in the majority of studies (e.g., where it was not possible to verify minimum duration of pain in the study sample, or the cut-off for successful outcome markedly differed from its common definition); (d) imprecision of the effect estimate in most studies, which could stem from inadequate sample size or lack of precision in reporting the effect size; and (e) potential publication bias assumed if the value of a

specific prognostic factor has not been repeatedly investigated (e.g. in ≥4 studies) or if smaller studies tended to report significant / larger effects relative to larger studies. In case of presence of factors increasing the confidence in available evidence, the quality ratings were upgraded for (f) moderate (d = 0.5, OR = 2.5, r = 0.3) or large (d = 0.8, OR = 4.25, r = 0.5) effect sizes; and (g) possible 'dose' effect within or between the studies, where higher levels of the prognostic factor lead to larger effect size.

3. Results

3.1. Study selection

The searches identified 2622 unique records. **Figure 1** illustrates the flow of the studies through the selection process. **Text S5, Supplemental Digital Content 5** includes a list of screened full text reports that were not eligible for the current review, with specific reasons for exclusion. Most articles were excluded due to ineligible population studied, that is, including patients with symptom duration <3 months or those with history of previous spine surgeries. For 5 reports [48–52] it was not possible to confirm whether all patients experienced symptoms for ≥3 months because the minimum duration data was not available, however, these reports were included in the review based on available average duration data suggesting chronic LBP. In total, 21 eligible reports of 18 studies assessing predictors of pain or disability outcomes were included in the review. Nine studies reported both outcomes [10,48–51,53–57], while 4 focused only on pain [9,11,52,58], and 5 only on disability [59–65].

3.2. Study characteristics

The characteristics of included studies are summarized in **Table 2**. There were 10 single- and 8 multi-center studies, spanning 13 different countries, with the United States, Sweden, and Switzerland being the most common locations. Ten studies had prospective and 8 retrospective design. Eight reports [10,11,51,53,60–62,64] were classified as phase-2 studies when considering specific independent associations whereas the remaining reports were classified as phase-1 studies. The

majority of included cohorts had spinal stenosis or disc herniation pathologies and underwent decompression or fusion surgeries. The follow-up duration ranged from 3 to 48 months after surgery, with a median of 14.5 across all studies. We included only the last available follow-up, unless eligible analysis was only conducted for an earlier time point [54].

Note that Gepstein et al. [48] analyzed two ethnic groups separately, which we consider to be 2 cohorts, whereas 2 reports from Kim et al. [49,50] appear to be based on 2 largely overlapping populations, thus we attribute these results to a single study cohort in this review. Three other reports [60–62] were classified as post-hoc subgroup analyses of one interventional study, and since each report conducted separate analyses on spinal stenosis and degenerative spondylolisthesis subgroups, these are considered as two unique cohorts in the current review.

3.3. Risk of bias in included studies

The 2 reviewers agreed on 83% of QUIPS domain ratings across 21 included reports, with Cohen's kappa = 0.72, 95% CI [0.61 to 0.82] indicating substantial agreement [66,67], before reaching 100% consensus. **Table S6, Supplemental Digital Content 6** presents the final domain-specific and overall RoB ratings for each included report, and **Figure 2** depicts the summary of these ratings across all reports.

Most studies were rated to have low RoB in outcome measurement, prognostic factor measurement, and study confounding domains. Despite a variety of measures used to assess the same predictors or outcomes, most studies used validated instruments and justified cut-offs. Serious study limitations (high RoB) in domains most relevant to prognosis were found in 2 studies for study participation [53,65] and in 9 studies for study attrition [10,49,50,52,54,56,60–62], but none for outcome measurement. High study participation RoB resulted from a lack of reporting regarding sample recruitment and insufficient details to confirm adequate participation of eligible subjects. Serious study limitations in the study attrition domain were most common and stemmed from inadequate response rates and no attempts to record or report reasons for and characteristics of

participants lost to follow-up to determine if there were important differences between those who completed the study and those who did not. Regarding other QUIPS domains, 2 studies were judged to have severe limitations in study confounding [49,57], and 4 in statistical analysis and reporting [48,56,57,63] domains. High study confounding RoB was related to unadjusted analysis or only partial reporting of the method of adjustment used and definition of exact confounders. Severe limitations of statistical analysis and reporting concerned inadequate sample size for the number of prognostic factors analyzed, incomplete reporting of the statistical models used (including underlying assumptions, e.g. multicollinearity), and indications of selective results reporting (e.g., lacking information on candidate predictors that were not statistically significant). Notably, the assessment of RoB relies on the level of reporting and in many cases it was downgraded due to unclear reporting or missing information. Overall, 14 reports (66%) were assessed to be at high RoB (with one or more domains judged as high) and 7 (33%) as low RoB (with all domains judged as low or moderate) [45].

3.4. Results of syntheses

In the interest of brevity, below we present detailed syntheses for factors that were found to be predictors of the outcomes of interest, whereas non-predictors are briefly summarized in the main text and more detailed results and discussion can be found in **Text S7, Supplemental Digital Content 7**.

3.4.1. Primary outcome: change in pain intensity

Fourteen included reports based on 13 studies examined predictors of pain relief in 14 patient cohorts (5780 participants in total). Most studies measured pain using 0-10 VAS [9,48–50,53] or NRS [10,54,58] scales, however, 0-100 VAS [11,51,52,56,57] and Pain Index [55] scales were also used. Ten studies assessed change in pain intensity as a continuous outcome [10,11,48–52,54–57], and 5 studies as a dichotomous one. The latter defined clinically significant improvement as 30% [53,58], 70% [9], 18/100 points [11], or 2/10 points [10] reduction in pain from preoperative baseline to

postoperative follow-up. Change in leg and back pain were examined as separate outcomes in 6 studies [10,11,49–51,56,57], 2 considered leg pain alone [52,54], and 5 assessed pain in general [9,48,53,55,58]. **Table 3** presents the results for each of these outcomes separately, however, since the within-study findings were largely consistent across back and leg pain outcomes, in the narrative synthesis we refer to change in pain intensity in general as a single outcome.

3.4.1.1. Sociodemographic predictors

Four studies investigated the associations between demographic and socioeconomic factors and pain outcomes in 5 cohorts including a total of 680 unique patients.

Three studies (1 low, 2 high RoB) investigated the association between *age* and pain outcomes, revealing inconclusive evidence. One study (high RoB) reported that older age was associated with less pain relief after surgery in adjusted analysis in an Arab cohort (moderate effect), and in unadjusted analyses in both Arab and Jewish cohorts (moderate and small effects) [48], whereas two studies found no association between these factors across adjusted and unadjusted analyses [9,53]. This inconsistency may stem from the fact that Gepstein et al. [48] recruited a significantly older population (≥65 years) compared to other studies.

A single high RoB study [48] examined the association of *education level* with pain outcomes in two ethnic cohorts, reporting better pain outcomes with increasing number of years of education in a Jewish cohort in adjusted and unadjusted analyses (small effects), but no significant association in an Arab cohort in unadjusted analysis. The Jewish cohort had significantly higher average education level compared to the Arab cohort (9.73 vs. 8.06 years), suggesting potential 'dose' effect within this study.

No significant associations with pain relief were found for other sociodemographic predictors, including *gender* (1 low, 3 high RoB studies [9,48,53,57]), *ethnicity* (1 high RoB study [48]), *work status* (1 phase-2 high RoB study [53]), or *workers compensation* (1 high RoB study [53]).

3.4.1.*2*. Health-related predictors

Eleven studies investigated the associations between health- and symptom-related factors and pain outcomes in 12 cohorts including a total of 5308 unique patients.

Four studies (2 low, 2 high RoB) [11,48,52,55] investigated whether type of *spinal pathology* was associated with pain outcomes. One phase-2 low RoB study reported greater reductions in back pain in patients with degenerative disc disease and herniated nucleus pulposus without (but not with) radiculopathy relative to those with spinal stenosis, and greater reductions in leg pain in patients with degenerative disc disease and herniated nucleus pulposus with radiculopathy, relative to spinal stenosis [11]. The second, phase-1 high RoB study reported greater reductions in leg pain in patients with multilevel (relative to single-level) stenosis and those with (relative to without) spondylolisthesis [52]. Analyses in both studies were adjusted for potential confounders, and defined change in pain levels as continuous outcomes. Unadjusted analyses in 2 low and 1 high RoB studies found no significant associations with type of pathology in an isthmic spondylolisthesis cohort [55] or in a Jewish cohort with spinal stenosis [48], except that not having *sciatica* was related to better outcomes in an Arab cohort with spinal stenosis (small effect) [48], and more patients with degenerative disc disease than with stenosis achieved significant reduction in back pain [11].

Four studies (1 low RoB, 3 high RoB) [9,10,54,56] investigated the prognostic effect of *symptom duration* on pain outcomes, overall showing evidence for no effect. One phase-2 high RoB study found no effect of 6-12 or >12 months duration of conservative treatment on reductions in leg and back pain in adjusted analyses [10]. Unadjusted analyses across three studies corroborated this finding, except for 1 high RoB study suggesting that patients with disc herniation who had symptoms for 3-12 months reported greater reductions in leg and back pain than those who had symptoms for >12 months [56]. The latter effect sizes were small; thus, we did not downgrade the evidence for symptom duration for inconsistency.

Two studies (1 low RoB, 1 high RoB) investigated the association between preoperative *pain intensity* and pain outcomes. Anderson [53] found that higher pain intensity was a marginally significant independent predictor of poorer pain outcomes (small effect), whereas Hegarty et al. [9] found no associations between these factors across adjusted and unadjusted analyses. This inconsistency is unlikely to be clinically relevant and may stem from the differences in the definition of outcome and duration of follow-up (30% pain reduction 24 months after surgery [53] vs. 70% pain reduction 3 months after surgery [9]). Average pain intensity was comparable across the two cohorts (6.8/10 [53] vs. 6.5 and 5.7/10 [9]).

One high RoB study assessed the predictive value of having *night-time pain*, reporting moderate adverse effect of this factor on pain reduction in unadjusted analyses in both ethnic cohorts [48].

A single high RoB study [48] assessed the effects of number, type and severity of *comorbidities* on pain outcomes in Arab and Jewish cohorts. In adjusted analyses, greater pain relief was predicted by absence of peripheral arterial disease and osteoarthrosis (small effects), absence of diabetes in the Arab cohort (moderate effect); and having lower number of comorbidities, not having peripheral arterial disease, diabetes, osteoarthritis, and no history of total joint replacement in the Jewish cohort (small effects). Unadjusted analyses showed that lower number of comorbidities and not having osteoarthritis (small effects), and absence of peripheral arterial disease and diabetes in the Arab cohort (moderate effects); and lower number of comorbidities, no history of total joint replacement, and lower American Society of Anesthesiologists class (small effects), not having peripheral arterial disease and diabetes (moderate effects), and absence of osteoarthritis in the Jewish cohort (large effect), were associated with more favorable pain outcomes. Lower American Society of Anesthesiologists class reflects patients' better physical status based on preoperative comorbid conditions.

The same high RoB study [48] tested the effect of *body mass index* on pain outcomes. Results of unadjusted analyses in Arab and Jewish cohorts indicate small effects of lower body mass index on greater pain reduction.

We found no significant associations with pain reduction for other investigated health-related factors, including *pain quality* (1 low RoB study [9]), *sensory detection and pain thresholds* (2 low RoB studies [9,58]), *conditioned pain modulation* (1 low RoB study [58]), *disability* (2 low RoB studies including 1 phase-2 study, 2 high RoB studies [9,51–53]), and *smoking* (1 high RoB study [53]).

3.4.1.3. Psychological predictors

Four studies (5 reports) investigated the associations between psychological factors and pain outcomes in 5 cohorts including a total of 746 unique patients.

Two studies (1 low RoB, 1 high RoB) evaluated the effect of *pain catastrophizing* on pain outcomes in unadjusted analyses. One low RoB study reported moderate effect size for lower total pain catastrophizing score, and significant effects of lower scores on the helplessness, rumination, and magnification subscales in relation to achieving ≥70% reduction in pain 3 months after surgery [9]. The high RoB study found the opposite effect (although it was not possible to estimate its size), whereby dichotomized high, compared to low, pain catastrophizing group reported greater reduction in back pain up until 12 months after surgery [50]. The latter study also reported no significant effect of pain catastrophizing on leg pain outcome [50].

One low RoB study tested the effect of *anxiety* on pain outcomes in unadjusted analyses [9]. Patients who achieved ≥70% reduction in pain had lower baseline levels on anxiety. While independent t-test analysis indicated a moderate effect, it did not replicate when Spearman's rank correlation was used.

Two studies (1 low RoB, 1 high RoB) examined the effect of *depression* on pain outcomes in unadjusted analyses. One high RoB study reported moderate effects of absence of depression

diagnosis on better pain outcomes in Arab and Jewish cohorts [48]. Another, low ROB study found no significant association between self-reported depression score and ≥70% pain reduction [9]. Inconsistent findings may arise from different operationalizations of the predictor, outcome, and the duration of follow-up.

No significant associations were found for the remaining psychological factors, including *pain sensitivity* (1 high RoB study [49]), *pain drawing* (1 low RoB study [55]), and *mental functioning* (low RoB study [9]).

3.4.2. Secondary outcome: change in disability

Seventeen included reports based on 14 studies examined predictors of disability outcomes in 15 patient cohorts (6899 participants in total). Several studies measured disability (or physical functioning) using ODI [49–51,54,56,59–62,64,65], SF-36 Physical Functioning (PF) subscale [49,50,57,60–62], SF-12 Physical Component Summary (PCS) [51,54], and RMDQ [53,63]. There were also single studies using COMI [10], PROMIS Physical Function (PF) subscale [51], Disability Rating Index [55], and Barthel Index [48]. Ten studies assessed change in disability as a continuous outcome [10,48–51,54–57,60–63], and 5 studies as a dichotomous one. The latter defined clinically significant improvement as 30% (RMDQ, ODI; [53,64]), 50% (ODI; [59]), 17/100 points (ODI; [65]), or 2/10 points (COMI; [10]) reduction in disability from preoperative baseline to postoperative follow-up. **Table 4** presents the results for each of the disability measures separately, however, since the within-study findings were largely consistent across different measures, in the narrative synthesis we refer to change in disability as a single outcome.

3.4.2.1. Sociodemographic predictors

Four studies investigated the associations between demographic and socioeconomic factors and disability outcomes in 4 cohorts including a total of 656 unique patients. None of the evaluated factors was found to be related to reduction in disability, including *age* (1 high RoB study [53]),

gender (3 high RoB studies [53,57,65]), *ethnicity* (1 high RoB study [48]), *work status* (1 high RoB phase-2 study [53]), and *workers compensation* (1 high RoB study [53]).

3.4.2.2. Health-related predictors

Seven studies (9 reports) investigated the associations between health- and symptom-related factors and disability outcomes in 8 cohorts including a total of 3715 unique patients.

Five studies (1 low, 4 high RoB) examined the prognostic value of disease *duration* for disability outcomes. One phase-2 study found that patients with spinal stenosis who had symptoms for <12 months reported greater improvement in physical function than those with longer symptom duration (small effect), but duration had no effect on disability in patients with degenerative spondylolisthesis in adjusted analyses [61]. However, another phase-2 study [10] and 1 phase-1 study [54] found no effects of disease duration on disability outcomes in spinal stenosis cohorts across adjusted and unadjusted analyses, resulting in mixed evidence regarding the prognostic value of symptom duration in this type of pathology. Two remaining studies provided inconsistent evidence in disc herniation cohorts, where one suggested that symptom duration <12 months was associated with greater reduction in disability (small effect) [56], and the other found no significant effect of disease duration [64].

A single phase-2 low RoB study [64] investigated whether *sensory detection threshold* is an independent predictor of disability outcomes, reporting that greater sensory loss (higher threshold) was associated with greater odds of achieving a clinically significant reduction in disability (moderate effect) across adjusted and unadjusted analyses. Out of multiple QST parameters measured in this study, the authors considered only sensory detection threshold as a candidate predictor of disability outcomes, as it significantly differed between patients and control participants.

Four studies (2 low, 2 high RoB) examined whether preoperative *disability* predicts disability outcomes. One phase-2 study (low RoB) and 1 phase-1 study (high RoB) found that less severe

baseline disability was an independent predictor of greater improvement postoperatively (unclear and moderate effects, respectively) across a range of continuous disability outcomes except change in ODI [51,63]. Two other studies (low and high RoB) reported no significant association between baseline disability and achieving MCID in disability in adjusted [53] and unadjusted [64] analyses. This inconsistency could not be explained by specific study characteristics, such as population, type of surgery, study design, or follow-up duration - there were no consistent differences between the studies reporting negative association and no association between baseline disability and surgery outcome.

We found no evidence for significant prognostic value of other health-related factors, including having *sciatica* (1 low RoB study [55]), *pain intensity, bothersomeness, and its neuropathic component* (1 low, 1 high RoB study [53,64]), *body mass index* (1 phase-2 low RoB study [60,62]), *smoking* (1 high RoB study [53]), and *sleep quality* (1 low RoB study [64]).

3.4.2.3. Psychological predictors

Four studies (5 reports) assessed the relationships between psychological factors and disability outcomes in 4 cohorts including 560 unique patients in total.

A single low RoB study [55] investigated the relationship between *pain drawing* and disability outcomes. Unadjusted analysis suggested that patients with organic pain drawing reported greater reduction in disability than those with non-organic pain drawing, however, the effect was only marginally significant and due to insufficient results reporting it was not possible to estimate its magnitude or precision.

Two studies (1 low, 1 high RoB) examined the effect of *depression* on change in disability. One unadjusted analysis suggested that lower depression scores were associated with greater improvement of disability [63], while the other one found no significant association [64]. Both studies examined similar disc herniation cohorts undergoing discectomy, yet the discrepancy in their

findings could be attributed to different measures of depression (Psychological general well-being index vs HADS) and disability (continuous RMDQ reduction vs dichotomous 30% reduction in ODI)

One high RoB study [63] assessed the association between *vitality* and disability outcomes, indicating that high vitality was related to greater improvement of disability in unadjusted analysis.

A single high RoB study [63] investigated the effect of *job-related resignation* on disability outcomes, demonstrating a moderate effect of lower resignation on greater improvement of disability in adjusted analysis, and consistent significant effect in unadjusted analysis.

One low RoB study [59] assessed the effect of *neuroticism* on change in disability. Unadjusted analysis suggested that lower neuroticism was associated with higher odds of achieving at least 50% reduction in ODI (small effect).

The remaining psychological factors, including *pain catastrophizing* (1 low, 1 high RoB study [50,64]), *pain sensitivity* (1 high RoB study [49]), *kinesiophobia* (1 low RoB study [64]), *mental functioning* (1 low RoB study [64]), and *anxiety* (1 low RoB study [64]) showed no significant associations with disability outcomes.

3.5. Quality of evidence

Detailed GRADE assessment of the quality of evidence is presented in **Supplemental Digital Content 8** (**Table S8a** for pain outcomes, **Table S8b** for disability outcomes) and summaries of findings regarding each outcome are presented in **Figure 3**. GRADE was carried out at the level of the following predictor categories: demographic factors (age, gender, ethnicity), socioeconomic characteristics (education, work status, worker's compensation), diagnosis (spinal pathology, sciatica), symptom duration, pain (intensity, quality, nigh-time pain, bothersomeness, neuropathic component of pain), quantitative sensory testing (sensory and pain thresholds, conditioned pain modulation), disability, comorbidities (comorbid conditions, body mass index, smoking), pain-related psychological factors (pain catastrophizing, pain sensitivity, pain drawing, kinesiophobia), affective-

motivational (mental functioning, anxiety, depression, vitality, job-related resignation), and personality factors (neuroticism).

Overall, we found moderate-quality evidence that *pain reduction* after surgery is not related to symptom duration, and low-quality evidence for no association with demographic or socioeconomic factors, although there was some indication that age may have a negative, and education level a positive effect on pain reduction, depending on population characteristics. We also found low-quality evidence that type of spinal pathology may be an independent predictor of pain relief, with absence of stenosis in particular being associated with more favorable outcomes. The evidence for the prognostic value of the remaining factor categories was of very low quality. Specifically, we found very low evidence for negative prognostic value of preoperative pain and comorbidities. There was also very low evidence for associations between psychological pain-related and affective factors with pain outcomes, with some indication of potential negative effects of pain catastrophizing, anxiety, and depression. Finally, there was very low evidence that preoperative disability and quantitative sensory testing are not independent predictors of pain reduction.

Most evidence contributing to the prediction of pain outcomes was from phase-1 studies, resulting in default moderate quality for all but 2 associations. More than half of the studies had high RoB, and in some low RoB studies, only unadjusted analyses were reported or eligible for the current review, resulting in downgrading the quality for severe study limitations in the majority of examined associations. Any inconsistencies in the results could be accounted for by the differences in study characteristics or measures used. In 4 out of 10 examined associations, the quality was downgraded for indirectness, mostly due to insufficient information in some studies to confirm that all patients had chronic pain [48–51]. Imprecision due to inadequate sample size or insufficient results reporting to enable evaluation of the precision of effect estimate (e.g. only reporting *p* values or omitting standard errors and confidence intervals for effect estimates) also contributed to downgrading the quality of evidence for the majority of associations [9,48,52,58]. Potential publication bias, which is a

common issue in prognostic factor research, was further exacerbated by the fact that many associations were evaluated in a very small number of studies. Only a few associations were supported by moderate / large effect sizes (for comorbidities and affective factors) or demonstrated potential 'dose' effects (demographic factors), thus there were limited basis for upgrading of confidence in the available evidence.

Regarding the secondary *disability* outcomes, there was moderate-quality evidence (from a single study) that greater sensory loss independently predicts more favorable disability outcomes. We found low-quality mixed evidence for the prognostic value of symptom duration, with some studies suggesting potential negative effect of longer disease duration on disability outcomes in patients with spinal stenosis or disc herniation, and others indicating no effect in the same pathologies as well as degenerative spondylolisthesis. Similarly, the effect of baseline disability on disability outcomes was unclear, with low-quality evidence indicating negative or no association. Furthermore, there was low-quality evidence suggesting that work-related socioeconomic factors and comorbidities (body mass index or smoking) are not independent predictors of disability outcomes. The evidence for the prognostic value of the remaining factor categories was of very low quality. Specifically, there was very low evidence that affective-motivational psychological factors (particularly job-related resignation) and neuroticism may be negatively related to improvement of disability. There was also very low-quality evidence that demographic factors, presence of sciatica, baseline pain features, sleep quality, and pain-related psychological factors do not predict disability outcomes.

Only 5 of the summarized associations were tested in confirmatory studies, while others were only explored in phase-1 studies, resulting in moderate starting quality for over half of the predictor categories. Most were downgraded for severe study limitations, as the majority of the available evidence was based on high RoB studies, with some relying only on unadjusted analyses. For half of the reported associations, inconsistency was not a problem as they were based on single studies.

The remaining predictor categories demonstrated consistent findings, except for baseline disability and symptom duration where the discrepancies could not be explained by any differences between the included studies. Only pain-related psychological factors were downgraded for indirectness, where studies with uncertainty regarding pain duration eligibility formed half of contributing evidence. The quality of evidence was downgraded for imprecision in over half of the summarized associations, mainly due to insufficient results reporting to evaluate the precision of reported effect estimates, and in fewer cases inadequate sample size [48,59,63–65]. All associations except for demographic factors, symptom duration, disability, and pain-related psychological factors were downgraded for potential publication bias because the available evidence was based on single or very few studies. The only cases for upgrading the level of evidence based on reported moderate effect size were sensory function and job-related resignation, however, for the latter the overall quality rating remained very low due to other serious concerns. There was no indication of 'dose' effects in any of the predictor categories.

4. Discussion

We systematically reviewed and synthesized the existing evidence regarding preoperative predictors of reduction in pain and disability after spinal surgery for chronic LBP and leg pain. The key findings are that for both outcomes, sociodemographic characteristics have overall limited prognostic value, and there is uncertain evidence for possible importance of psychological factors. Among the health-related factors, there is a potential effect of type of spinal pathology, and less certain effects of preoperative pain and comorbidities on the primary pain outcome, whereas sensory loss is likely associated with the secondary disability outcome, and the evidence is mixed regarding potential effects of symptom duration and preoperative disability on the same outcome.

This review advances the existing literature by summarizing a range of potential predictors of pain intensity and disability outcomes, which are highly important to patients and constitute the most common surgical goals [10,68]. Operationalizing these outcomes as changes from baseline to follow-

up allowed for more precise quantification of reduction in pain and improvement in disability. Our synthesis further benefits from the thorough assessment of quality of evidence, both at the study level and the level of particular associations, using tools specifically adapted to prognostic studies [44,47].

4.1. Sociodemographic predictors

Although sociodemographic factors appear to be unrelated to pain or disability outcomes (low-quality evidence), we found some evidence suggesting that lower general education level and older age might be associated with less pain relief from surgery, depending on sample characteristics. Specifically, a positive effect of education on pain reduction may only manifest in patients with higher education levels [48]. Similarly, presence of both significant (in older spinal stenosis cohorts [48]) and not significant (in younger discogenic LBP cohorts) effects of age may suggest that this association manifests only in older age groups. Three previous reviews reported similar associations between older age and worse surgery outcomes [15,28], or younger age and better outcomes [19], in populations largely affected by disc herniation. The prognostic effect of age might therefore depend on the type of spinal pathology.

4.2. Health-related predictors

Several health-related factors were identified as potential predictors of reduction in pain or disability after surgery. Previous systematic reviews concerned with this class of predictors focused only on disc herniation cohorts [15,18,19], thus our review extends the existing evidence synthesis to a broader range of spinal pathologies. In fact, the type of diagnosis itself may be an important predictor, as cohorts with spinal stenosis presented with less reduction in back and leg pain after surgery compared to degenerative disc disease and disc herniation cohorts (low-quality evidence). Another study (not eligible for this review) concluded that differences in outcomes may not depend on the specific pathology, but rather on patients' age as those with spinal stenosis tend to be older than patients with disc herniation [69]. The impact of these two factors may be difficult to disentangle.

Contrary to our expectation that patients with chronic LBP (implying longer symptom duration) could experience less benefit from surgery, we found moderate-quality evidence that symptom duration is likely unrelated to pain outcomes, while its effect on reduction in disability appears to be mixed (low-quality evidence). Similar inconsistencies are apparent among previous reviews indicating either no association [18] or negative effect of longer symptom duration on spinal surgery outcomes [19,28]. The mixed evidence may depend on the composition of the studied samples: studies in which most patients experienced symptoms for <12 months tended to report greater benefit from earlier surgery [56,61], and those in which most patients had symptoms for >12 months reported no significant associations with surgery outcomes [10,54]. While 12 months cut-off was commonly used to distinguish longer and shorter symptom duration, it is possible that a lower cut-off, e.g., the point at which LBP becomes chronic, would allow better discrimination between favorable and unfavorable outcomes.

In line with two previous reviews [15,28], we found very low-quality evidence that less preoperative pain may be an independent predictor of better pain (but not disability) outcomes, although this effect appears sensitive to the pain reduction cut-off used. There was moderate-quality evidence for the association between greater sensory loss in the affected extremity (consistent with nerve root compression) and more favorable disability outcomes, suggesting greater improvement in the context of clear neurological pathology that can be directly addressed by surgery.

The current inconsistent evidence concerning the effect of baseline disability on pain relief (very low-quality) weights towards lack of association, whereas the evidence for disability outcomes (low-quality) could not be easily reconciled, similar to previous reviews [15,18,28]. Possibly, worse baseline disability may be related to smaller improvement in disability after surgery as a continuous outcome [51,63], but not as MCID [53,64].

Medical comorbidities are known to increase the risk of postoperative complications [70,71], however, their effect on longer-term spinal surgery outcomes is less clear [19,72]. We found very

low-quality evidence supporting significant associations between present comorbidities and smaller benefit from surgery in terms of reduction in pain, but not disability (low-quality).

4.3. Psychological predictors

We found several significant associations between psychological factors and pain and disability outcomes, but based on very low-quality evidence. Previous systematic reviews identified depression, anxiety, somatization, neuroticism, poor coping, and catastrophizing as important predictors of spinal surgery outcomes [15,22,28,30] and our findings in chronic LBP cohorts are largely in agreement. Lower pain catastrophizing and anxiety appear to be related to improved pain-specific but not disability outcomes. Negative effects of depression were also present, but not consistently for either outcome, potentially due to varying assessment methods. Furthermore, we found a significant independent association between lower job-related resignation and greater improvement in disability, which may be related to higher motivation to return to work after surgery. Finally, neuroticism, which reflects a predisposition to experience negative affect and maladaptive responses to stress and is considered a risk factor for a range of health problems [73], was also found to predict less improvement in disability after spinal surgery.

4.4. Limitations

Our confidence in the reviewed associations is limited by the quality of available evidence, as detailed in sections 3.3 and 3.5. The overall low / very low quality stems from the dominance of exploratory rather than confirmatory studies, lack of adjustment for potential confounders in several associations, and imprecision of effect estimates related to insufficient results reporting or inadequate sample size. Another reason is high overall RoB in 66% of the included studies, with severe limitations most prevalent in the study attrition and analysis and reporting domains. Therefore, some of the reported relationships are likely to be different for cases with and without complete follow-up, inadequately representing the studied samples, and some of the reported results are likely to be spurious or biased due to inadequate analysis or reporting [44]. Confidence in independent prognostic value of certain preoperative factors is also limited by different sets of

potential confounders used across the reviewed studies, as the magnitude and significance of any associations may depend on other included predictors. The quality of evidence is further limited by the fact that many associations were examined only in a small number of studies. This may be because certain potential predictors, such as psychological risk factors, are rarely formally documented before spinal surgery, e.g. in prospective spine registries [74–76]. Considering high exclusion rate due to ineligible symptom duration, it is possible that including studies with unrestricted duration would provide additional evidence for the reviewed associations in LBP more generally, although this would be beyond the scope of the current review concerning chronic LBP populations.

Quantitative synthesis of the results of reviewed studies was not possible due to their methodological heterogeneity and incomplete reporting, and some decisions regarding the review process may have contributed to this. We included several different measures of pain and disability outcomes, as these were commonly used and validated in the population of interest. We further considered both dichotomous and continuous outcomes due to their clinical utility and precision, respectively. Although included follow-up intervals covered a broad range of stages of recovery (3-48 months), patient-reported outcomes following spinal surgery assessed over multiple time periods are known to be strongly correlated [77,78]. The scope of the review included a range of spinal pathologies and types of surgery under the assumption that there are prognostic factors that are common across different populations and interventions. While this approach has further contributed to the heterogeneity of results, at the same time it strengthens the generalizability of identified associations, especially between health-related factors and surgery outcomes, which have been previously summarized only in specific patient populations [15,18,19].

To use all available data, some effect measures had to be estimated based on reported results, extracted from figures, or recalculated for eligible subgroups, which could have added uncertainty to the results synthesis. We also included both adjusted and unadjusted analyses where available.

Although unadjusted effect estimates provided lower-quality evidence due to potential alternative explanatory factors, they can uncover predictors of interest worthy of further investigation.

Our search strategy did not seek non-English language or 'grey' literature. Although these restrictions motivated by pragmatic reasons and limited resources may have introduced a potential information bias, there is no consistent evidence that language-restricted reviews lead to biased effect estimates [79], and including unpublished or non-peer-reviewed sources of evidence could have limited the precision and confidence in the results. Finally, while data extraction, QUIPS, and GRADE assessments were conducted dually, the reviewers' decisions were not entirely independent, as each reviewer primarily assessed half of the included studies and verified the other reviewer's judgements for the remaining studies before joint discussions to reach consensus.

4.5. Implications

Through comprehensive evaluation of the existing evidence regarding preoperative predictors of reduction in pain and disability after spinal surgery for chronic LBP, we have highlighted certain gaps and issues that should be addressed by further research. Prospective studies could address the participation and attrition biases by reporting the characteristics of patients who were excluded as non-eligible or lost to follow-up. Power analysis and transparent reporting of the analytic assumptions and methods used, and provision of complete data including non-significant results would prevent potentially spurious results contributing to low quality of evidence. Confirmatory studies testing the direction and strength of independent associations while controlling for potential confounders are also needed to provide higher certainty in evidence regarding prognosis [47]. We propose that factors which presented low or very low-quality evidence of potential significant (baseline pain, comorbidities, and psychological factors) or unclear relationships with spinal surgery outcomes (symptom duration and baseline disability) should be tested as independent predictors, while sensory deficits and type of spinal pathology (possibly interacting with age) are adjusted for. If the prognostic value of these factors is confirmed in future studies, further research would be

warranted to evaluate the effectiveness of interventions addressing the modifiable risk factors before surgery to improve its outcomes.

4.6. Conclusions

The success of spinal surgery for chronic LBP is susceptible to clinical heterogeneity of patients. We found a likely association between sensory loss and improved disability outcomes, and a potential relationship of spinal stenosis with less pain relief. Age and general education may also contribute to the extent of pain reduction, depending on population characteristics. While these predictors could potentially assist in weighing risks and benefits when deciding on the best course of treatment, at the current quality of evidence they should not determine qualification for surgery. Other sociodemographic factors do not appear to predict surgery outcomes, while symptom duration is likely unrelated to pain outcomes but may be adversely related to disability outcomes, similar to baseline disability. The associations between spinal surgery outcomes and other potential predictors are less certain. More high-quality confirmatory studies are needed to establish reliable prognostic factors for patients with chronic LBP.

Acknowledgements

We are grateful to Dr Sarah Nevitt for her insightful comments on the systematic review protocol.

Supplemental Digital Content

Supplemental Digital Content 1.doc Search strategies for electronic databases

Supplemental Digital Content 2.doc Elaboration on eligibility criteria and decision rules

Supplemental Digital Content 3.doc Data extraction form template

Supplemental Digital Content 4.doc Risk of bias assessment form template

Supplemental Digital Content 5.doc Excluded full text reports with reasons

Supplemental Digital Content 6.doc Risk of bias judgements

Supplemental Digital Content 7.doc Results of syntheses and discussion of non-predictors of pain

and disability outcomes

Supplemental Digital Content 8.doc GRADE quality of evidence assessment

References

1. Hoy D, Bain C, Williams G, March L, Brooks P, Blyth F, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum 2012;64(6):2028–37.
2. Costa L da CM, Maher CG, Hancock MJ, McAuley JH, Herbert RD, Costa LOP. The prognosis of acute and persistent low-back pain: a meta-analysis. CMAJ 2012;184(11):E613–24.
3. Hayden JA, Dunn KM, van der Windt DA, Shaw WS. What is the prognosis of back pain? Best Pract Res Clin Rheumatol 2010;24(2):167–79.
4. Rice AS, Smith BH, Blyth FM. Pain and the global burden of disease. Pain 2016;157(4):791–6.
5. Manchikanti L, Singh V, Falco FJ, Benyamin RM, Hirsch JA. Epidemiology of low back pain in adults. Neuromodulation Technol Neural Interface 2014;17:3–10.
6. Rajaee SS, Bae HW, Kanim LEA, Delamarter RB. Spinal fusion in the United States: analysis of trends from 1998 to 2008. Spine 2012;37(1):67–76.
7. Weir S, Samnaliev M, Kuo T-C, Choitir CN, Tierney TS, Cumming D, et al. The incidence and healthcare costs of persistent postoperative pain following lumbar spine surgery in the UK: a cohort study using the Clinical Practice Research Datalink (CPRD) and Hospital Episode Statistics (HES). BMJ Open 2017;7(9):e017585.
8. Abbott TEF, Fowler AJ, Dobbs TD, Harrison EM, Gillies MA, Pearse RM. Frequency of surgical treatment and related hospital procedures in the UK: a national ecological study using hospital episode statistics. BJA Br J Anaesth 2017;119(2):249–57.
9. Hegarty D, Shorten G. Multivariate prognostic modeling of persistent pain following lumbar discectomy. Pain Physician 2012;15(5):421–34.
10. Zweig T, Enke J, Mannion AF, Sobottke R, Melloh M, Freeman BJ, et al. Is the duration of pre-operative conservative treatment associated with the clinical outcome following surgical decompression for lumbar spinal stenosis? A study based on the Spine Tango Registry. Eur Spine J 2016/12/17 edn 2017;26(2):488–500.

11. Zweig T, Hemmeler C, Aghayev E, Melloh M, Etter C, Röder C. Influence of preoperative nucleus pulposus status and radiculopathy on outcomes in mono-segmental lumbar total disc replacement: results from a nationwide registry. BMC Musculoskelet Disord 2011;12(1):275–275.
12. de Campos TF. Low back pain and sciatica in over 16s: assessment and management NICE Guideline [NG59]. J Physiother 2017;63(2):120.
13. Kreiner DS, Hwang S, Easa J, Resnick DK, Baisden J, Bess S, et al. Diagnosis and Treatment of Lumbar Disc Herniation with Radiculopathy. North American Spine Society; 2012100. (Evidence-Based Clinical Guidelines for Multidisciplinary Spine Care).
14. Rolving N, Sogaard R, Nielsen CV, Christensen FB, Bünger C, Oestergaard LG. Preoperative cognitive-behavioral patient education versus standard care for lumbar spinal fusion patients: economic evaluation alongside a randomized controlled trial. Spine 2016;41(1):18–25.
15. Dorow M, Lobner M, Stein J, Konnopka A, Meisel HJ, Gunther L, et al. Risk Factors for Postoperative Pain Intensity in Patients Undergoing Lumbar Disc Surgery: A Systematic Review. PLoS One 2017/01/21 edn 2017;12(1):e0170303.
16. Huysmans E, Goudman L, Van Belleghem G, De Jaeger M, Moens M, Nijs J, et al. Return to work following surgery for lumbar radiculopathy: a systematic review. Spine J 2018/05/26 edn 2018;18(9):1694–714.
17. Oosterhuis T, Smaardijk VR, Kuijer PPF, Langendam MW, Frings-Dresen MHW, Hoving JL. Systematic review of prognostic factors for work participation in patients with sciatica. Occup Env Med 2019/07/13 edn 2019;76(10):772–9.
18. Rushton A, Zoulas K, Powell A, Staal JB. Physical prognostic factors predicting outcome following lumbar discectomy surgery: systematic review and narrative synthesis. BMC Musculoskelet Disord 2018/09/13 edn 2018;19(1):326.

19. Wilson CA, Roffey DM, Chow D, Alkherayf F, Wai EK. A systematic review of preoperative predictors for postoperative clinical outcomes following lumbar discectomy. Spine J 2016/08/09 edn 2016;16(11):1413–22.

20. Daubs MD, Norvell DC, McGuire R, Molinari R, Hermsmeyer JT, Fourney DR, et al. Fusion versus nonoperative care for chronic low back pain: do psychological factors affect outcomes? Spine Phila Pa 1976 2011/10/05 edn 2011;36(21 Suppl):S96-109.

21. Mroz TE, Norvell DC, Ecker E, Gruenberg M, Dailey A, Brodke DS. Fusion versus nonoperative management for chronic low back pain: do sociodemographic factors affect outcome? Spine Phila Pa 1976 2011/10/05 edn 2011;36(21 Suppl):S75-86.

22. Wilhelm M, Reiman M, Goode A, Richardson W, Brown C, Vaughn D, et al. Psychological Predictors of Outcomes with Lumbar Spinal Fusion: A Systematic Literature Review. Physiother Res Int 2015/08/14 edn 2017;22(2).

23. Vasseljen O, Woodhouse A, Bjrngaard JH, Leivseth L. Natural course of acute neck and low back pain in the general population: The HUNT study. PAIN 2013;154(8):1237–44.

24. Dhondt E, Van Oosterwijck J, Cagnie B, Adnan R, Schouppe S, Van Akeleyen J, et al. Predicting treatment adherence and outcome to outpatient multimodal rehabilitation in chronic low back pain. J Back Musculoskelet Rehabil 2019/07/30 edn 2020;33(2):277–93.

25. O'Donnell JA, Anderson JT, Haas AR, Percy R, Woods ST, Ahn UM, et al. Preoperative Opioid Use is a Predictor of Poor Return to Work in Workers' Compensation Patients After Lumbar Diskectomy. Spine 03622436 2018;43(8):594–602.

26. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in healthcare. University of York: CRD; 2009.

27. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.

28. 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 2009/07/30 edn 2009;10(4):639–53.
29. Gaudin D, Krafcik BM, Mansour TR, Alnemari A. Considerations in Spinal Fusion Surgery for Chronic Lumbar Pain: Psychosocial Factors, Rating Scales, and Perioperative Patient Education-A Review of the Literature. World Neurosurg 2016/11/05 edn 2017;98:21–7.
30. McKillop AB, Carroll LJ, Battie MC. Depression as a prognostic factor of lumbar spinal stenosis: a systematic review. Spine J 2014/01/15 edn 2014;14(5):837–46.
31. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. J Med Libr Assoc JMLA 2018;106(4):531.
32. Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ 2019/02/01 edn 2019;364:k4597.
33. Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LO, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. Pain 2018;159(3):481.
34. Pincus T, Santos R, Breen A, Burton AK, Underwood M. A review and proposal for a core set of factors for prospective cohorts in low back pain: a consensus statement. Arthritis Care Res Off J Am Coll Rheumatol 2008;59(1):14–24.
35. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980/08/01 edn 1980;66(8):271–3.
36. Roland M, Morris R. A study of the natural history of back pain: Part 1: Development of a reliable and sensitive measure of disability in low-back pain. spine 1983;
37. Deyo RA, Battie M, Beurskens A, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research: a proposal for standardized use. Spine 1998;23(18):2003–13.
38. Ghogawala Z, Resnick DK, Watters WC, Mummaneni PV, Dailey AT, Choudhri TF, et al. Guideline update for the performance of fusion procedures for degenerative disease of the

lumbar spine. Part 2: assessment of functional outcome following lumbar fusion. J Neurosurg Spine 2014;21(1):7–13.

39. Mannion AF, Elfering A, Staerkle R, Junge A, Grob D, Semmer NK, et al. Outcome assessment in low back pain: how low can you go? Eur Spine J 2005;14(10):1014–26.
40. Effect Size Calculator [Internet]. [cited 2021 Feb 23]. Available from: https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php
41. Lenhard W, Lenhard A. Computation of Effect Sizes [Internet]. 2017.
42. Grooten WJA, Tseli E, Ang BO, Boersma K, Stalnacke BM, Gerdle B, et al. Elaborating on the assessment of the risk of bias in prognostic studies in pain rehabilitation using QUIPS-aspects of interrater agreement. Diagn Progn Res 2019/05/17 edn 2019;3:5.
43. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006;144(6):427–37.
44. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med 2013;158(4):280–6.
45. Bruls VEJ, Bastiaenen CHG, de Bie RA. Prognostic factors of complaints of arm, neck, and/or shoulder: a systematic review of prospective cohort studies. Pain 2015/02/07 edn 2015;156(5):765–88.
46. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol 2011;64(4):380–2.
47. Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. Syst Rev 2013;2(1):1–12.
48. Gepstein R, Arinzon Z, Folman Y, Shabat S, Adunsky A. Lumbar spine surgery in Israeli Arabs and Jews: a comparative study with emphasis on pain perception. Isr Med Assoc J IMAJ 2007;9(6):443–7.

49. Kim HJ, Lee JI, Kang KT, Chang BS, Lee CK, Ruscheweyh R, et al. Influence of pain sensitivity on surgical outcomes after lumbar spine surgery in patients with lumbar spinal stenosis. Spine Phila Pa 1976 2014/11/11 edn 2015;40(3):193–200.

50. Kim HJ, Park JW, Chang BS, Lee CK, Yeom JS. The influence of catastrophising on treatment outcomes after surgery for lumbar spinal stenosis. Bone Jt J 2015;97-B(11):1546-1554.

51. Patel DV, Bawa MS, Haws BE, Khechen B, Block AM, Karmarkar SS, et al. PROMIS Physical Function for prediction of postoperative pain, narcotics consumption, and patient-reported outcomes following minimally invasive transforaminal lumbar interbody fusion. J Neurosurg Spine 2019;30(4):476–82.

52. Sigmundsson FG, Kang XP, Jonsson B, Stromqvist B. Prognostic factors in lumbar spinal stenosis surgery. Acta Orthop 2012/10/23 edn 2012;83(5):536–42.

53. Anderson PA, Schwaegler PE, Cizek D, Leverson G. Work status as a predictor of surgical outcome of discogenic low back pain. Spine 2006;31(21):2510–5.

54. Cushnie D, Thomas K, Jacobs WB, Cho RKH, Soroceanu A, Ahn H, et al. Effect of preoperative symptom duration on outcome in lumbar spinal stenosis: a Canadian Spine Outcomes and Research Network registry study. Spine J 2019;19(9):1470–7.

55. Ekman P, Moller H, Hedlund R. Predictive factors for the outcome of fusion in adult isthmic spondylolisthesis. Spine 2009;34(11):1204–10.

56. Støttrup CC, Andresen AK, Carreon L, Andersen MØ. Increasing reoperation rates and inferior outcome with prolonged symptom duration in lumbar disc herniation surgery — a prospective cohort study. Spine J 2019;19(9):1463–9.

57. Stromqvist F, Ahmad M, Hildingsson C, Jonsson B, Stromqvist B. Gender differences in lumbar disc herniation surgery. Acta Orthop 2008/10/08 edn 2008;79(5):643–9.

58. Muller M, Limacher A, Agten CA, Treichel F, Heini P, Seidel U, et al. Can quantitative sensory tests predict failed back surgery?: A prospective cohort study. Eur J Anaesthesiol 2019;36(9):695–704.

59. Hägg O, Fritzell P, Ekselius L, Nordwall A. Predictors of outcome in fusion surgery for chronic low back pain. A report from the Swedish Lumbar Spine Study. Eur Spine J 2003;12(1):22-33.

60. McGuire KJ, Khaleel MA, Rihn JA, Lurie JD, Zhao W, Weinstein JN. The effect of high obesity on outcomes of treatment for lumbar spinal conditions: subgroup analysis of the spine patient outcomes research trial. Spine Phila Pa 1976 2014/11/05 edn 2014;39(23):1975–80.

61. Radcliff KE, Rihn J, Hilibrand A, DiIorio T, Tosteson T, Lurie JD, et al. Does the duration of symptoms in patients with spinal stenosis and degenerative spondylolisthesis affect outcomes?: analysis of the Spine Outcomes Research Trial. Spine Phila Pa 1976 2011/09/14 edn 2011;36(25):2197–210.

62. Rihn JA, Radcliff K, Hilibrand AS, Anderson DT, Zhao W, Lurie J, et al. Does obesity affect outcomes of treatment for lumbar stenosis and degenerative spondylolisthesis? Analysis of the Spine Patient Outcomes Research Trial (SPORT). Spine Phila Pa 1976 2012/05/23 edn 2012;37(23):1933–46.

63. Schade V, Semmer N, Main CJ, Hora J, Boos N. The impact of clinical, morphological, psychosocial and work-related factors on the outcome of lumbar discectomy. Pain 1999;80(1–2):239–49.

64. Tampin B, Slater H, Jacques A, Lind CRP. Association of quantitative sensory testing parameters with clinical outcome in patients with lumbar radiculopathy undergoing microdiscectomy. Eur J Pain 2020;24(7):1377–92.

65. Watkins RG, O'Brien JP, Draugelis R, Jones D. Comparisons of preoperative and postoperative MMPI data in chronic back patients. Spine 1986;11(4):385–90.

66. Altman DG. Practical statistics for medical research. CRC press; 1990.

67. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37–46.

68. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations: Pain 2003;106(3):337–45.

69. Herron LD, Turner J, Clancy S, Weiner P. The differential utility of the minnesota multiphasic personality inventory: A predictor of outcome in lumbar laminectomy for disc herniation versus spinal stenosis. Spine 1986;11(8):847–50.

70. Bays A, Stieger A, Held U, Hofer LJ, Rasmussen-Barr E, Brunner F, et al. The influence of comorbidities on the treatment outcome in symptomatic lumbar spinal stenosis: A systematic review and meta-analysis. North Am Spine Soc J NASSJ 2021;6:100072.

71. Luo W, Sun R, Jiang H, Ma X. The effect of diabetes on perioperative complications following spinal surgery: a meta-analysis. Ther Clin Risk Manag 2018;14:2415–23.

72. Jiang J, Teng Y, Fan Z, Khan S, Xia Y. Does Obesity Affect the Surgical Outcome and Complication Rates of Spinal Surgery? A Meta-analysis. Clin Orthop Relat Res 2014;472(3):968–75.

73. Lahey BB. Public Health Significance of Neuroticism. Am Psychol 2009;64(4):241–56.

74. Asher AL, Knightly J, Mummaneni PV, Alvi MA, McGirt MJ, Yolcu YU, et al. Quality Outcomes Database Spine Care Project 2012–2020: milestones achieved in a collaborative North American outcomes registry to advance value-based spine care and evolution to the American Spine Registry. Neurosurg Focus 2020;48(5):E2.

75. Melloh M, Staub L, Aghayev E, Zweig T, Barz T, Theis J-C, et al. The international spine registry SPINE TANGO: status quo and first results. Eur Spine J 2008;17(9):1201.

76. Strömqvist B, Fritzell P, Hägg O, Jönsson B, Sandén B, Swedish Society of Spinal Surgeons. Swespine: the Swedish spine register : the 2012 report. Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc 2013;22(4):953–74.

77. Parker SL, Asher AL, Godil SS, Devin CJ, McGirt MJ. Patient-reported outcomes 3 months after spine surgery: is it an accurate predictor of 12-month outcome in real-world registry platforms? Neurosurg Focus 2015/12/02 edn 2015;39(6):E17.

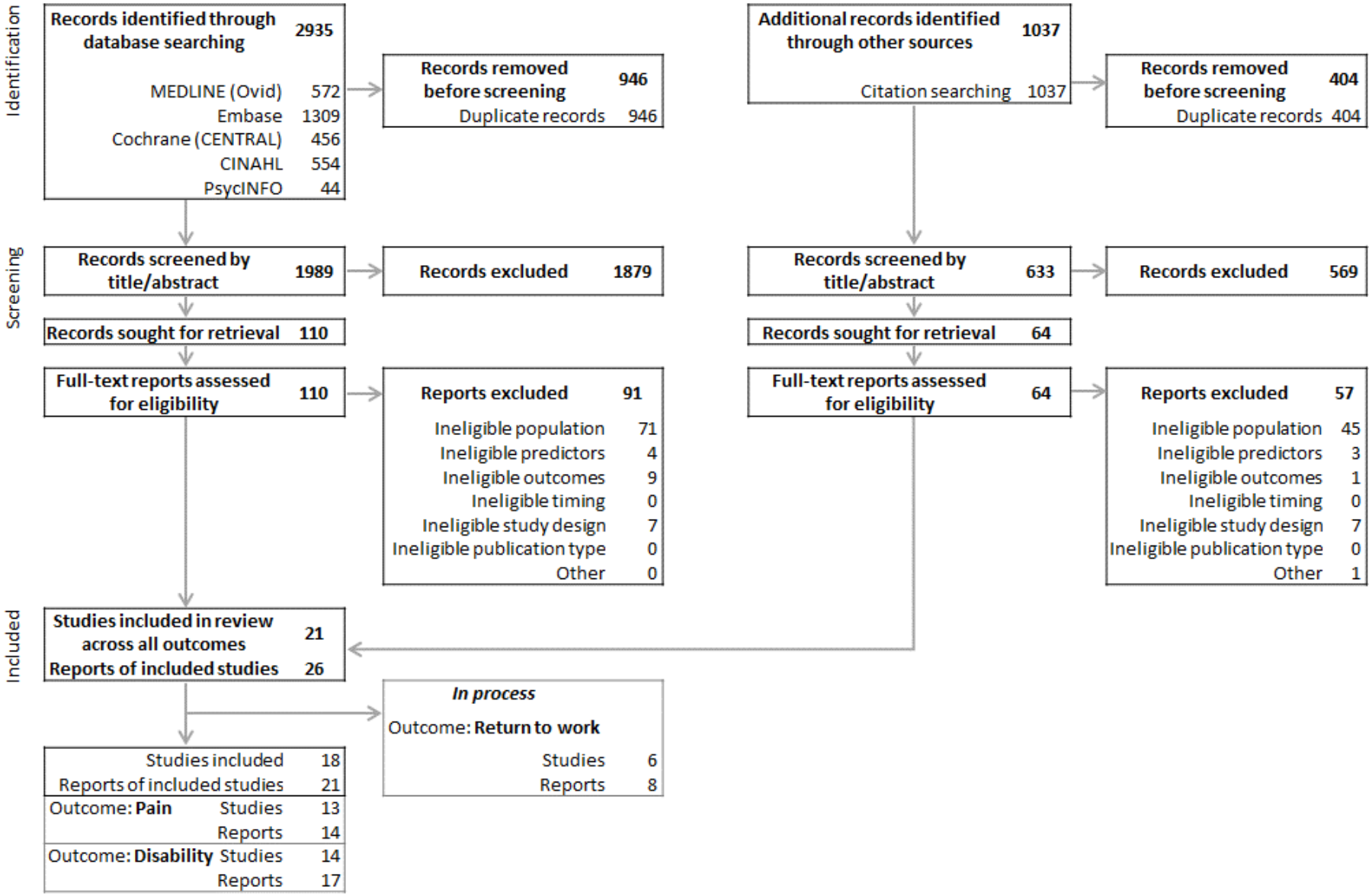
78. Staartjes VE, Siccoli A, de Wispelaere MP, Schröder ML. Patient-reported outcomes unbiased by length of follow-up after lumbar degenerative spine surgery: Do we need 2 years of follow-up? Spine J Off J North Am Spine Soc 2019;19(4):637–44.
79. Moher D, Pham, Klassen TP, Schulz KF, Berlin JA, Jadad AR, et al. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol 2000;53(9):964–72.

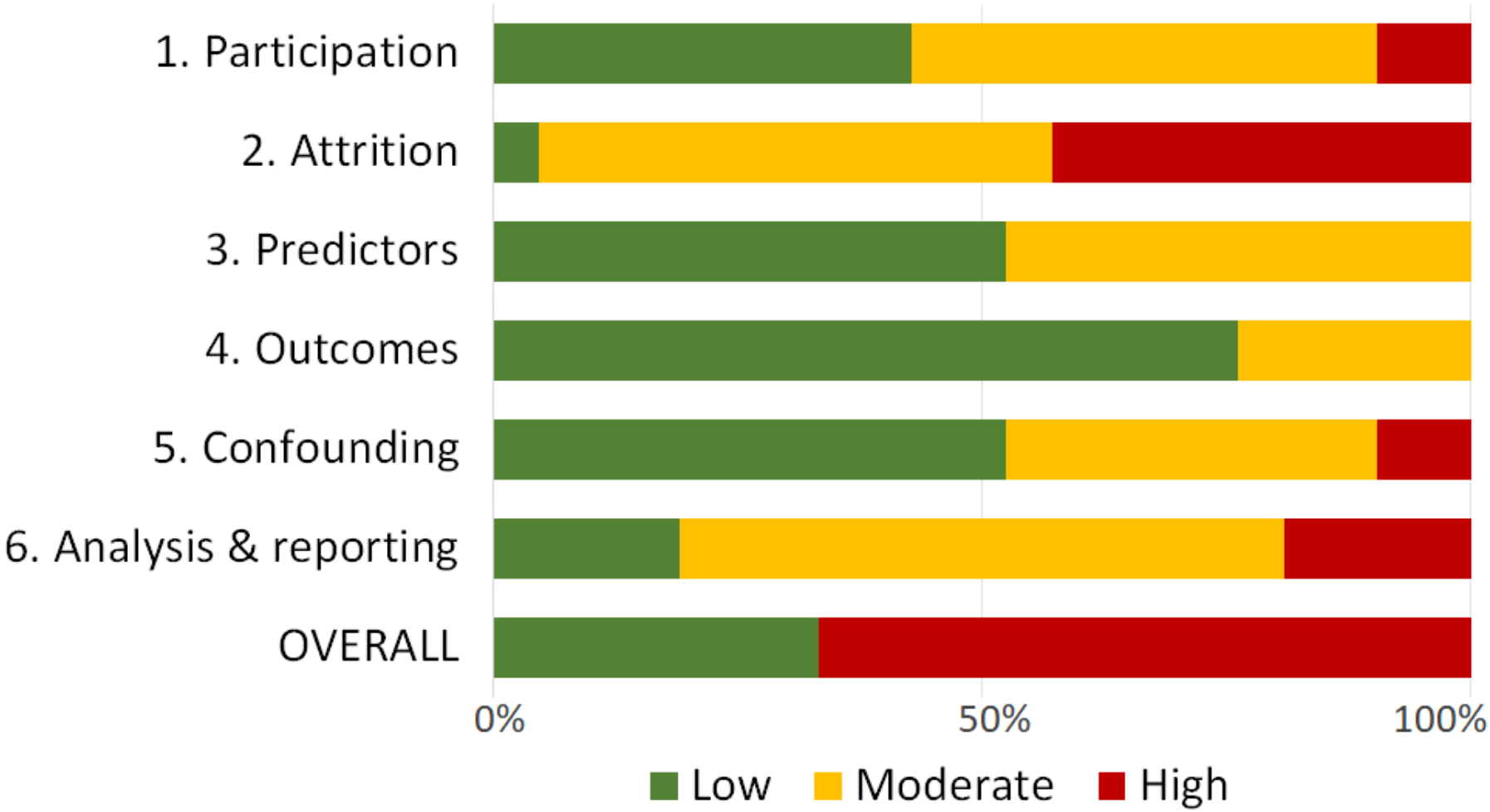
Figure legends

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram [27]. Reports reflect individual published articles, and in some cases more than one report based on the same study was published. Note that some studies reported more than one outcome of interest, therefore, the numbers of included studies and reports per outcome do not add up to the total number of included studies and reports.

Figure 2. Risk of bias across 21 included reports in each Quality of Prognosis Studies (QUIPS) [44] domain and overall assessment of risk of bias across all domains.

Figure 3. Overall quality of evidence for the reviewed associations according to Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework [46,47].
QST, quantitative sensory testing.


Identification
Records identified through database searching 2935
MEDLINE (Ovid) 572
Embase 1309
Cochrane (CENTRAL) 456
CINAHL 554
PsycINFO 44
Records removed before screening 946
Duplicate records 946
Additional records identified through other sources 1037
Citation searching 1037
Records removed before screening 404
Duplicate records 404
Screening
Records screened by title/abstract 1989
Records excluded 1879
Records screened by title/abstract 633
Records excluded 569
Records sought for retrieval 110
Records sought for retrieval 64
Full-text reports assessed for eligibility 110
Reports excluded 91
Ineligible population 71
Ineligible predictors 4
Ineligible outcomes 9
Ineligible timing 0
Ineligible study design 7
Ineligible publication type 0
Other 0
Full-text reports assessed for eligibility 64
Reports excluded 57
Ineligible population 45
Ineligible predictors 3
Ineligible outcomes 1
Ineligible timing 0
Ineligible study design 7
Ineligible publication type 0
Other 1
Included
Studies included in review across all outcomes 21
Reports of included studies 26
In process
Outcome: Return to work
Studies 6
Reports 8
Studies included 18
Reports of included studies 21
Outcome: Pain Studies 13
Reports 14
Outcome: Disability Studies 14
Reports 17


1. Participation
2. Attrition
3. Predictors
4. Outcomes
5. Confounding
6. Analysis & reporting
OVERALL
0%
50%
100%
Low
Moderate
High

	Change in pain intensity			Change in disability		
Baseline predictors	*N of cohorts (patients)*	*Effect*	*Overall quality*	*N of cohorts (patients)*	*Effect*	*Overall quality*
Sociodemographic						
Demographic	5 (680)	✗	●●○○	4 (656)	✗	●○○○
Socioeconomic	3 (326)	✗	●●○○	1 (93)	✗	●●○○
Health-related						
Diagnosis	5 (893)	✓	●●○○	1 (164)	✗	●○○○
Symptom duration	4 (4066)	✗	●●●○	5 (4474)	?	●●○○
Pain	4 (379)	✓	●○○○	2 (141)	✗	●○○○
QST	2 (116)	✗	●○○○	1 (48)	✓	●●●○
Disability	4 (318)	✗	●○○○	4 (313)	?	●●○○
Comorbidities	3 (326)	✓	●○○○	3 (897)	✗	●●○○
Sleep				1 (48)	✗	●○○○
Psychological						
Pain-related	3 (388)	✓	●○○○	3 (383)	✗	●○○○
Affective-motivational	3 (273)	✓	●○○○	2 (90)	✓	●○○○
Personality				1 (183)	✓	●○○○

Sig. association:
Yes ✓, No ✗,
Unclear ?

Quality of evidence:

●●●● High
●●●○ Moderate
●●○○ Low
●○○○ Very low

Table 1. Inclusion and exclusion criteria.

	Include	Exclude
Population	• Adults (≥18 years) • Chronic (lasting or recurring for ≥3 months) low back pain and / or lumbar radicular pain (pain radiating to the leg due to nerve root compression) • Primary lumbar / lumbosacral spine surgery	• Revision surgery / history of previous lumbar spine surgery • Pathology of tumor, trauma, infection, or inflammatory disease • Spinal Cord Stimulator implantation / injections, chemical or radiofrequency interventions
Predictors	• Preoperative assessment of prognostic factors	• Intraoperative, genetic, or radiographic predictors
Outcomes	• Change in back and / or leg pain intensity • Change in function / disability • Change in Core Outcome Measure Index	• Only postoperative assessment of outcomes without baseline reference • Pain / disability assessed only using measures without a continuous score or as part of a composite outcome
Timing	• Outcomes assessed ≥3 months post-surgery (no upper follow-up limit)	
Setting	• Spinal surgery sites or registries / databases of operated patients	
Study design	• Randomized / nonrandomized controlled study • Cohort study • Case-control study • Registry / database study	• No investigation of associations between preoperative factors and postoperative outcomes • Case study / series
Publication type	• Original research • Peer-reviewed • English language • Publication period from 1984	• Review • Conference abstract

Table 2. Study characteristics.

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
Anderson 2006 [53]; Orthopedic Surgery and Rehabilitation Department, University Hospital (US)	Prospective cohort	Exclusion: significant psychosocial abnormalities following psychological assessment	N = 106; Age <48 and >48; Duration ≥6 months	Discogenic LBP	Anterior lumbar interbody fusion	Work status at time of surgery; Smoking; Gender; Worker's compensation; Age; Baseline pain; Baseline disability; Levels fused; Cage type	Change in pain (0-10 VAS); Change in disability (RMDQ)	24 months (81%)
Cushnie 2019 [54]; 18 Neurosurgery and Orthopedic Spine Surgery Hospitals, part of	Retrospective cohort (registry-based)	Inclusion: neurogenic claudication or radiculopathy as chief complaint; Exclusion:	N = 466; Age M = 65, SD = 11; 62% male; Duration 6% 3-6, 18% 6-12, 27% 12-24, 46% >24 months	Degenerative stenosis	Decompression	Symptom duration	Change in disability (ODI); Change in physical functioning (SF-12 PCS); Change in leg pain (0-10 NRS)	12 months (69%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
prospective multicenter Canadian Spine Outcomes and Research Network registry (Canada)		scoliosis, spondylolisthesis						
Ekman 2009 [55]; University Hospital, Spine Centre, 2 General Hospitals (Sweden)	Prospective cohort	Inclusion: severely restricted function for ≥1 year; Exclusion: drug or alcohol abuse, psychiatric disorders	N = 164; Age 18–55, M = 40; 43% male; Duration ≥12 months	Isthmic spondylolisthesis	47% posterolateral fusion (78% 1-level, 22% multi-level), 53% posterolateral interbody fusion	Pain drawing; Sciatica; Work status[d]	Change in pain (Pain Index); Change in disability (DRI)	24 months (98%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
					(76% 1-level, 24% multi-level)			
Gepstein 2007 [48]; 2 Spinal Units, University Hospital (Israel)	Retrospective case-control (internal registry)	Inclusion: disabling back/leg pain and progressive decline in walking ability	N = 220; [Arab cohort] N = 69; Age 65-NR, M = 71; 53% male; Duration M = 58.08 months; [Jewish cohort] N = 151; Age 65-NR, M = 72; 52% male; Duration M = 42.61 months	Spinal stenosis	58% decompressive laminectomy, 22% discectomy, 20% both	Ethnicity; Leg pain; Night-time pain; Peripheral arterial disease; Diabetes; Osteoarthritis; Total joint replacement; Depression; Number of comorbidities; ASA class; Gender; Age; BMI; Education level	Change in pain (VAS); Change in function (Barthel Index)	M = 46 months (80%)
Hagg 2003 [59]; 19 Orthopedic	RCT, post-hoc analysis	Inclusion: severe CLBP of ≥2 years	N = 201; Age 25–64, M = 43, SD =	Degenerative spondylosis	Posterolateral fusion (67%	Neuroticism	Change in disability (ODI)	24 months (91%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
Departments (Sweden)		duration, back pain more severe than leg pain, no radiculopathy, ≥1 year of sick leave / equivalent disability / failed conservative treatment, ≥7/10 Function and Working Disability Score; Exclusion: psychiatric illness, spondylolisthesis, spinal stenosis, painful and	8; 49% male; Duration 24–408, M = 94.08, SD = 81.84 months		instrumented, 33% non-instrumented)			

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
		disabling arthritic hip joints						
Hegarty 2012 [9]; Neurosurgery Institute (Ireland)	Prospective cohort	Inclusion: ASA classification I-II, failed ≥12 weeks non-operative treatment; Exclusion: cauda equina syndrome, spinal or genetic abnormalities, pregnancy, gabapentin / pregabalin / opioids use 2	N = 53; [PPSP] N = 20; Age 27–50, Mdn = 40; 45% male; Duration 3-60, Mdn = 9; [no PPSP] N = 33; Age 22–55, Mdn = 39; 57% male; Duration 3-48, Mdn = 6 months	Disc herniation with nerve root compression	Microdiscectomy (open)	Age; Gender; Duration of pain; Pain quality and severity; Physical disability; Anxiety; Depression; Pain coping strategies; Health-related quality of life; Sensory, pain perception, and tolerance thresholds	Persistent post-surgical pain (change in pain on movement, 0-10 VAS)	3 months (100%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
		weeks before surgery, other pain syndrome occurring ≥weekly with ≥3/10 intensity, refusal to return for follow-up						
Kim 2015a[c] [49]; Orthopedic Surgery Department, University Hospital (South Korea)	Prospective cohort	Exclusion: history of major psychiatric disorder, peripheral vascular disease, concurrent	N = 171; Age 40–80, M = 62, SD = 14; 36% male; Duration M = 10.91, SD = 9.05 months	Spinal stenosis	40% fusion, 60% decompression	Pain sensitivity	Change in disability (ODI); Change in back pain (0-10 VAS); Change in leg pain (0-10 VAS); Change in physical	12 months (73%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
		serious medical condition					functioning (SF-36 PCS)	
Kim 2015b[c] [50]; Orthopedic Surgery Department, University Hospital (South Korea)	Prospective cohort	Exclusion: history of major psychiatric disorder, peripheral vascular disease, concurrent serious medical condition	N = 138; Age 40–80, M = 66, SD = 11; 34% male; Duration M = 11.89, SD = 8.91 months	Spinal stenosis	48% fusion, 52% decompression; 70% 1-level, 30% multi-level	Pain catastrophizing	Change in disability (ODI); Change in back pain (0-10 VAS); Change in leg pain (0-10 VAS); Change in physical functioning (SF-36 PCS)	12 months (75%)
McGuire 2014[c] [60]; 13 Multidisciplinary Spine Clinics (US)	RCT & observational cohort, post-	Inclusion: neurogenic claudication or radicular pain	N = 413; Age M = 63, SD = 12; 62% male; Duration	Spinal stenosis	87% decompression, 6% instrumented fusion, 5% non-	Extreme obesity; Age; Gender; Race; Smoking status; Compensation status; Comorbidities	Change in physical function (SF-36 PF); Change in disability	48 months (70%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
	hoc subgroup analysis	with neurologic signs, symptoms for ≥12 weeks, surgical candidates; Exclusion: spondylosis, isthmic spondylolisthesis	61% <12, 39% ≥12 months N = 391; Age M = 65, SD = 10; 31% male; Duration 65% <12, 35% ≥12 months	Degenerative spondylolisthesis	instrumented fusion 71% instrumented fusion, 21% non-instrumented fusion, 6% decompression	(joint, stomach, bowel and intestinal problems, osteoporosis, other); Number of moderately/severely stenotic levels; Self-assessed baseline health trend; Treatment preference; Baseline stenosis bothersomeness; Baseline score on outcome measure; Centre	(ODI, MODEMS version)	
Muller 2019 [58]; 3 tertiary care	Prospective cohort	Inclusion: lumbosacral	N = 141; Age M = 61, SD = 14; 42%	Degenerative pathology: 55%	Decompression (35% without	Electrical pain detection thresholds; Pressure pain	Failed back surgery syndrome	12 months (97%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
centers, Department of Anesthesiology and Pain Medicine (Switzerland)		radiculopathy, chronic low back pain ≥3/10 on NRS on most days/week; Exclusion: bilateral pain below the knees, neurological comorbidities, psychiatric comorbidities (except depression), previous instrumented	male; Duration 24% >60, 100% ≥3 months	spinal stenosis, 63% spondylolisthesis, 77% endplate changes, 15% scoliosis, 49% severe facet joint degeneration, 87% severe disc degeneration, 12% ≥50% fatty degeneration muscles	instrumental stabilization); 68% 1-level, 32% multi-level	detection and tolerance thresholds; Heat pain detection thresholds; Cold pain detection thresholds; Cold pressor test; Conditioned pain modulation; Type of surgery; Number of operated segments; Gender; Catastrophizing; BMI; Lasègue sign; Finger-floor distance; Baseline disability; Non-opioid and opioid analgesics intake	(persistence of pain, 0-10 NRS)	

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
		spinal surgery, current surgery >3 segments, multiple somatic comorbidities, unable to contact before surgery						
Patel 2019 [51]; Orthopedic Surgery Department (US)	Retrospective cohort (internal registry)	Inclusion: surgery between 2015-2017; Exclusion: multi-level fusion, unavailable pre-operative PROMIS or immediate post-operative	N = 130; Age M = 52, SD = 11; 58% male; Duration M = 33.2, SD = 50, Mdn = 14 months	Degenerative pathology	1-level transforaminal interbody fusion	Baseline disability; BMI; Worker's compensation	Change in physical function (PROMIS PF); Change in disability (ODI); Change in physical functioning (SF12 PCS); Change in back pain (VAS);	12 months (100%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
		pain and narcotics use data					Change in leg pain (VAS)	
Radcliff 2011[c] [61]; 13 Multidisciplinary Spine Clinics (US)	RCT & observational cohort, post-hoc subgroup analysis	Inclusion: neurogenic claudication or radicular pain with neurologic signs, symptoms for ≥12 weeks, surgical candidates; Exclusion: spondylosis, isthmic spondylolisthesis	N = 413; Age M = 63, SD = 12; 62% male; Duration 61% <12, 39% ≥12 months	Spinal stenosis	86% decompression, 6% instrumented fusion, 5% non-instrumented fusion	Symptom duration; Age; Gender; BMI; Race; Smoking status; Compensation status; Comorbidities (joint, stomach, bowel and intestinal problems, osteoporosis, other); Number of moderately / severely stenotic levels; Self-assessed baseline health trend; Treatment preference; Baseline	Change in physical function (SF-36 PF); Change in disability (ODI, MODEMS version)	48 months (70%)
			N = 391; Age M = 65, SD = 10; 31% male; Duration 65% <12, 35% ≥12 months	Degenerative spondylolisthesis	71% instrumented fusion, 21% non-instrumented fusion, 5% decompression			

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
						stenosis bothersomeness; Baseline score on outcome measure; Centre		
Rihn 2012[c] [62]; 13 Multidisciplinary Spine Clinics (US)	RCT & observational cohort, post-hoc subgroup analysis	Inclusion: neurogenic claudication or radicular pain with neurologic signs, symptoms for ≥12 weeks, surgical candidates; Exclusion: spondylosis,	N = 413; Age M = 63, SD = 12; 62% male; Duration 61% <12, 39% ≥12 months N = 391; Age M = 65, SD = 10; 31% male; Duration 65% <12, 35% ≥12 months	Spinal stenosis Degenerative spondylolisthesis	87% decompression, 6% instrumented fusion, 5% non-instrumented fusion 71% instrumented fusion, 21% non-instrumented	Obesity; Age; Gender; Race; Smoking status; Compensation status; Comorbidities (joint, stomach, bowel and intestinal problems, osteoporosis, other); Number of moderately / severely stenotic levels; Self-assessed baseline health trend; Treatment	Change in physical function (SF-36 PF); Change in disability (ODI, MODEMS version)	48 months (70%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
		isthmic spondylolisthesis			fusion, 6% decompression	preference; Baseline stenosis bothersomeness; Baseline score on outcome measure; Centre		
Schade 1999 [63]; 3 Orthopedic and Neurosurgery Departments (Switzerland, Canada, US)	Prospective case-control, post-hoc subgroup analysis	Inclusion: employed, 6-8 weeks of failed conservative treatment, availability for clinical and MRI examination before surgery; Exclusion: rapid	N = 46; Age 20–50, M = 35; 74% male; Duration 46% 3-6, 26% 6-12, 28% >12 months	Disc herniation with radicular leg pain	Discectomy	Baseline disability; Extent of neural compromise; Job-related resignation; Depression; Vitality[e]	Change in disability (RMDQ)	24, range 23–30 months (91%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
		progressive motor deficit, cauda equina syndrome, no Swiss residency						
Sigmundsson 2012 [52]; Orthopedic Department, part of prospective Swedish Spine Register (Sweden)	Prospective cohort (internal registry)	Exclusion: fusion, high-grade spondylolisthesis, low-grade spondylolisthesis with spondylosis, instability, higher level of back pain than leg pain	N = 109; Age M = 71, SD = 10; 51% male; Duration 42% (leg) and 49% (back) >24 months	Spinal stenosis	Decompression	Walking distance; Spinal pathology; Age; Baseline leg and back pain; Duration of leg and back pain	Change in leg pain (0-100 VAS)	12 months (90%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
Støttrup 2019 [56]; Centre for Spine Surgery and Research, part of prospective national DaneSpine registry (Denmark)	Retrospective cohort (registry-based)	Inclusion: 1st episode, 12 weeks of failed conservative treatment; Exclusion: cauda equina syndrome, severe neurologic deficits	N = 1531; Age M = 46, SD = 15; 52% male; Duration 72% 3-12, 28% >12 months	Disc herniation with radicular leg pain	Discectomy	Duration of leg pain	Change in disability (ODI); Change in back pain (0-100 VAS); Change in leg pain (0-100 VAS)	12 months (79%)
Stromqvist 2008 [57]; Orthopedic Department, part of prospective Swedish Spine Register (Sweden)	Retrospective cohort (internal registry)		N = 301; Age 18–82, M = 42; 55% male; Duration most patients 3–12, 10% ≥24 months	Disc herniation	Disc degeneration surgery (41% microscopic, 59% open)	Gender	Change in leg pain (0-100 VAS); Change in back pain (0-100 VAS); Change in health outcomes (SF-36)	12 months (80%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
Tampin 2020 [64]; Neurosurgery Department (Australia)	Prospective cohort	Inclusion: radicular leg pain in L5 or S1 dermatomal distribution due to nerve root compression at L4/5 or L5/S1, elective neurosurgery waitlist; Exclusion: diabetes, vascular, neurological, or	N = 53; Age 18–65, M = 38, SD = 11; 51% male; Duration 4–50, M = 11.7, SD = 7.5 months	Disc herniation with radicular leg pain	Microdiscectomy	Mechanical detection threshold; Gender; Baseline leg and back pain intensity and bothersomeness; Baseline disability; Symptom duration; Anxiety; Depression; Pain catastrophizing; Kinesiophobia; Sleep quality; Physical and mental functioning	Change in disability (ODI)	12 months (91%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
		psychiatric disease						
Watkins 1986 [65]; Orthopedic Hospital (UK)	Retrospective cohort (internal registry)	Inclusion: disabling CLBP, available baseline and >2 years follow-up outcomes	N = 42; Age 27–56, M = 42, SD = 1; 55% male; Duration 12–300, M = 103.2, SD = 13.2 months	CLBP of intervertebral and nerve root etiology	Anterior interbody fusion	Gender; Age[d]	Change in disability (ODI)	≥24 months (100%)
Zweig 2011 [11]; SWISSspine prospective multicenter registry (Switzerland)	Retrospective cohort (registry-based)	Inclusion: ≥6 months failed conservative treatment, 1-3 years follow-up available	N = 433; Age 19-65, M = 42, SD = 9.2; 40% male; Duration ≥6 months	37% degenerative disc disease, 16% disc herniation without radiculopathy, 17% radiculopathy	1-level total disc replacement	Type of pathology; Gender; Age; Pre-operative pain medication; Intervertebral level operated; Depression;	Change in back pain (0-100 VAS); Change in leg pain (0-100 VAS)	12-35, M = 22, SD = 8 months (100%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
				without disc herniation (stenosis), 30% disc herniation with radiculopathy		Type of work; Working activity level		
Zweig 2017 [10]; Spine Tango prospective multicenter registry (Australia, Belgium, Germany, Poland,	Retrospective cohort (registry-based)	Inclusion: preoperative and ≥1 postoperative COMI, ASA classification, eligible for ≥3 months follow-up; Exclusion: spondylolisthesis,	N = 2016; Age 22-97, M = 68, SD = 11; 53% male; Duration 38% 6–12, 62% >12 months	Degenerative spinal stenosis	Decompression	Duration of conservative treatment; Age; Gender; ASA; Number of affected segments; Level operated; Surgical goal; Patient-reported main problem; Type of surgery; Follow-up duration;	Change in back pain (0-10 NRS); Change in leg pain (0-10 NRS); Change in COMI (0-10)	3–30, M = 17, SD = 8 months (100%)

Study ID, setting	Study type	Population				Index & Comparator prognostic factors[a]	Outcomes[a]	Timing
		Inclusion / exclusion criteria[b]	*Sample characteristics*	*Diagnosis (pathology)*	*Surgery (type, levels)*			*Follow-up (resp. rate)*
Switzerland, UK, US)		deformity, countries without a validated language COMI				Baseline back pain, leg pain, COMI score		

ASA, American Society of Anesthesiologists; BMI, Body Mass Index; CLBP, Chronic Low Back Pain; COMI, Core Outcome Measures Index; DRI, Disability Rating Index; HNP, herniated nucleus pulposus; LBP, Low Back Pain; M, mean; Mdn, median; MODEMS, Musculoskeletal outcomes Data Evaluation and Management Systems; MRI, Magnetic Resonance Imaging; NR, not reported; NRS, Numeric Rating Scale; ODI, Oswestry Disability Index; PCS, Pain Catastrophizing Scale; PPSP, Persistent Post-Surgical Pain; PROMIS, Patient-Reported Outcomes Measurement Information System; PSQ, Pain Sensitivity Questionnaire; RCT, Randomized Controlled Trial; RMDQ, Roland Morris Disability Questionnaire; SD, standard deviation; SF-36 / SF-12, Short Form Health Survey; VAS, Visual Analog Scale.

[a] Only eligible index prognostic factors or those included in eligible analyses as comparator prognostic factors are listed; only eligible outcomes are listed.

[b] Listed only inclusion/exclusion criteria that were not covered by the eligibility criteria for the systematic review.

[c] Reports based on the same or overlapping populations: Kim 2015a and Kim 2015b (largely overlapping cohorts); McGuire 2014, Radcliff 2011, and Rihn 2012 (the same two subgroups from the SPORT trial, spinal stenosis and degenerative spondylolisthesis cohorts are reported separately due to how they were analyzed).

[d] Data could not be extracted due to insufficient reporting.

[e] Unclear which additional candidate predictors were considered due to selection of factors for multivariate modeling based on significance of their univariate associations with outcome and missing information on those that were not significant.

Table 3. Effects of prognostic factors on change in pain from baseline to the last available follow-up.

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition*[a]	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors*[b] */ unadjusted*	*Estimate (95% CI)*[c]
	Sociodemographic					
	Age					
Anderson 2006 [53]	Age <48 years vs >48 years	VAS (0-10) 30% pain reduction	24 (M = 30)	106	I. Multivariate logistic regression; adjusted for: baseline work status (working), smoking (yes), gender (NR), worker's compensation (yes), age >48, baseline pain (0-10 VAS), baseline disability (RMDQ), levels fused (single), cage type (BAK)	I. <48 years: OR = 0.63 (0.21 to 1.92), p = .41; >48 years: OR = 0.78 (0.19 to 3.23), p = .73
Hegarty 2012 [9]	Age (years)	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53	I. Multivariate logistic backward stepwise regression; adjusted for: present pain intensity, disability (RMDQ) II. Mann-Whitney U test (unadjusted); III. Spearman's rho correlation (unadjusted)	I. b = -.05, SE = .04, Wald = 1.8, p = .16, OR = 1.0, chi-square = 1.9, p = .16 II. No PPSP Mdn = 39, range 22 – 55, PPSP Mdn = 40, range 27 – 50; U = 284, z = 0.85, p = .39; III. rho = -.11, p = .41;

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Gepstein 2007 [48]	Age (years)	VAS (0-10) pain change	M = 46	220 (69 Arab, 151 Jewish)	I. Multiple linear regression; adjusted for: [Arab] diabetes (yes), osteoarthritis (yes), peripheral arterial disease (yes) II. Spearman's rank correlation (unadjusted)	I. Arab: b = -0.27, SE = 0.08, beta = -0.31, t = -3.42, p = .001 (R^2 = 0.27) II. Arab: r = -.408, p < .001; Jewish: r = -.203, p < .05
	Gender					
Anderson 2006 [53]	NR (reference level not specified – direction of the effect uncertain)	VAS (0-10) 30% pain reduction	24 (M = 30)	106	I. Multivariate logistic regression; adjusted for: baseline work status (working), smoking (yes), worker's compensation (yes), age <48, age >48, baseline pain (0-10 VAS), baseline disability (RMDQ), levels fused (single), cage type (BAK)	I. OR = 1.46 (0.54 to 3.97), p = .45
Stromqvist 2008 [57]	Female vs male	VAS (0-100) leg pain mean reduction VAS (0-100) back pain mean reduction	12	301 (136 female, 165 male)	I. Analysis of covariance (unadjusted – covariates not specified)	I. Female M = 20, male M = 16, p > .05 I. Female M = 36, male M = 40, p > .05

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Gepstein 2007 [48]	Female vs male	VAS (0-10) pain change	M = 46	220 (69 Arab, 151 Jewish)	I. Wilcoxon rank sum test (unadjusted)	I. Arab: r = .27, p < .05; Jewish: r = .19, p < .05
Hegarty 2012 [9]	Male vs female	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53 (28 male, 25 female)	I. Chi-square test (unadjusted); II. Spearman's rho correlation (unadjusted)	I. 19 male, 14 female had no PPSP, * OR = 1.66 (0.54 to 5.08); II. rho = -.08, p = .56
	Ethnicity					
Gepstein 2007 [48]	Israeli Arabs vs Israeli Jews	VAS (0-10) pain mean reduction	M = 46	220 (69 Arab, 151 Jewish)	I. Independent samples t-test (unadjusted)	I. Arab M = 4.91, SD = 0.41, Jewish M = 4.85, SD = 2.7, p > .05, *d = -0.03 (-0.31 to 0.26)
	Education					

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Gepstein 2007 [48]	Years of education	VAS (0-10) pain change	M = 46	220 (69 Arab, 151 Jewish)	I. Multiple linear regression; adjusted for: [Jewish] diabetes (yes), osteoarthritis (yes), total joint replacement (yes), comorbidities (number); II. Spearman's rank correlation (unadjusted)	I. Jewish: b = 0.13, SE = 0.05, beta = 0.16, t = 2.53, p = .012 (R^2 = 0.44); II. Arab: r = .17, p > .05; Jewish: r = .27, p < .001
	Work status					
Anderson 2006 [53]	Working (including home working and studies) vs not working	VAS (0-10) 30% pain reduction	24 (M = 30)	105 (49 working, 65 not working)	I. Multivariate logistic regression; adjusted for: smoking (yes), gender (NR), worker's compensation (yes), age <48, age >48, baseline pain (0-10 VAS), baseline disability (RMDQ), levels fused (single), cage type (BAK)	I. OR = 0.78 (0.28 to 2.21), p = .64
	Worker's compensation					
Anderson 2006 [53]	Compensation claim vs no compensation	VAS (0-10) 30% pain reduction	24 (M = 30)	106 (50 compensation, 36 no compensation)	I. Multivariate logistic regression; adjusted for: baseline work status (working), smoking (yes), gender (NR), age <48, age >48, baseline pain (0-10 VAS), baseline disability (RMDQ), levels fused (single), cage type (BAK)	I. OR = 2.07 (0.75 to 5.75), p = .16

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	Health-related					
	Spinal pathology					
Zweig 2011 [11]	Degenerative disc disease (DDD) vs HNP without radiculopathy (HNP-NoRad) vs Stenosis vs HNP with radiculopathy (HNP-Rad)	VAS (0-100) back pain reduction	12-36 (M = 22)	433 (160 DDD, 68 HNP-NoRad, 73 stenosis, 132 HNP-Rad)	I. General linear model; adjusted for: gender (female); age; pre-operative pain medication (yes); intervertebral level operated (L3/4 / L4/5 / L5/S1); pharmacologically treated depression (yes); type of work (sedentary / physical / housewife / retired / unemployed); working activity level (unable to work / 10-40% / 50-90% / 100%)	I. DDD M = 49.8, HNP-NoRad M = 45.9, Stenosis M = 32.6, HNP-Rad = 45.2; Stenosis < DDD, p = .001; Stenosis < HNP-NoRad, p = .032; Stenosis < HNP-Rad, p = .064
		VAS (0-100) leg pain reduction				I. DDD M = 34.7, HNP-NoRad M = 33.4, Stenosis M = 21.8, HNP-Rad M = 34; Stenosis < DDD, p = .026; Stenosis < HNP-Rad, p = .040
		VAS (0-100) back pain ≥18 vs <18 points reduction			I. Univariate logistic regression (unadjusted)	I. 84% DDD, 77.4% HNP-NoRad, 60.6% Stenosis, 71.7% HNP-Rad ≥18 reduction; Stenosis < DDD, p = .002; Stenosis <

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
						HNP-NoRad, p = .16; Stenosis < HNP-Rad, p = .54
		VAS (0-100) leg pain ≥18 vs <18 points reduction				I. 66.7% DDD, 56.5% HNP-NoRad, 60.1% Stenosis, 71.7% HNP-Rad ≥18 reduction; HNP-Rad > DDD, p = 1.00; HNP-Rad > HNP-NoRad, p = .25; HNP-Rad > Stenosis, p = .77
Sigmundsson 2012 [52]	Levels of stenosis: multi-level vs 1-level	VAS (0-100) leg pain reduction	12	76	I. Multivariable regression; adjusted for: age, baseline leg and back pain (0-100 VAS), baseline walking distance, duration of leg and back pain (>24 months), spondylolisthesis (yes)	I. Multi-level: B = -15 (-30 to -0.2), p = .05
	Spondylolisthesis vs no spondylolisthesis				I. Multivariable regression; adjusted for: age, baseline leg and back pain (0-100 VAS), baseline walking distance, duration of leg and back pain (>24 months), stenosis (multilevel)	I. Spondylolisthesis: B = -16 (-31 to -1), p = .04
	Sciatica					

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Ekman 2009 [55]	Sciatica (pain symbols below the knee on Pain Drawing) vs no sciatica	Pain Index (0-100) mean reduction	24	164 (119 sciatica, 45 no sciatica)	I. Mann-Whitney U test (unadjusted)	I. Sciatica M = 19, no sciatica M = 30, p = .85
Gepstein 2007 [48]	Leg pain vs no leg pain	VAS (0-10) pain change	M = 46	220 (69 Arab, 151 Jewish)	I. Wilcoxon rank sum test (unadjusted)	I. Arab: r = -.26, p < .05; Jewish: r = -.05, p > .05
	Symptom duration					
Cushnie 2019 [54]	Duration of the main neurologic leg complaint (months): 3-6 vs 6-12 vs 12-24 vs >24	NRS (0-10) leg pain mean change	12	466 (26 3-6 months, 85 6-12 months, 125 12-24 months, 230 >24 months)	I. Wilcoxon signed-rank test (unadjusted)	I. 3-6 months M = -4.7 (-3.4 to -5.9); 6-12 months M = -4.5 (-3.8 to -5.3); 12-24 months M = -4 (-3.5 to -4.6); >24 months M = -3.7 (-3.3 to -4.2) [fig] *overlapping CIs - no significant differences*

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Zweig 2017 [10]	Duration of pre-operative conservative treatment (months): 6-12 vs >12	NRS (0-10) back pain ≥2 vs <2 points change	3 – 30 (M = 17)	2016 (758 6-12 months, 1258 >12 months)	I. Chi-square test after propensity score weighing adjustment for: age, gender (female), ASA (1 / 2 / >2), number of affected segments (1 / 2–3 / >3), level operated (L1/2–L2/3 / L3/4 / L4/5 / L5/S1), surgical goal (pain reduction / functional / neurological improvement), patient-reported main problem (back / leg pain / sensory disturbances / other), type of surgery, follow-up duration, baseline back pain, leg pain, COMI score; II. Multiple logistic regression; adjusted for inverse probability of treatment weight (propensity score), sequestrectomy (yes), foraminotomy (yes) III. Chi-square test (unadjusted);	I. 400 6-12 months, 715 >12 months had ≥2 change; *OR = 0.85 (0.71 to 1.02); II. OR = 0.85 (0.69 to 1.02) [fig]; III. 402 6-12 months, 720 >12 months had ≥2 change; *OR = 0.84 (0.70 to 1.01)

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
		NRS (0-10) leg pain ≥2 vs <2 points change				I. 510 6-12 months, 886 >12 months had ≥2 change; *OR = 0.86 (0.71 to 1.05); II. OR = 0.87 (0.72 to 1.05) [fig]; III. 512 6-12 months, 860 >12 months had ≥2 change; *OR = 0.96 (0.79 to 1.17)
		NRS (0-10) back pain mean reduction			I. General linear model after propensity score weighing adjustment for: age, gender (female), ASA (1 / 2 / >2), number of affected segments (1 / 2–3 / >3), level operated (L1/2–L2/3 / L3/4 / L4/5 / L5/S1), surgical goal (pain reduction / functional / neurological improvement), patient-reported main problem (back / leg pain / sensory disturbances /	I. 6-12 months M = 2.1, SD = 3.2, >12 months M = 2.2, SD = 3.2; *MD = 0.1 (-0.19 to 0.39), SE = 0.15, p = .497, d = 0.03 (-0.06 to 0.12); II. 6-12 months M = 2.1, SD = 3.2, >12 months M = 2.2, SD = 3.2; *MD = 0.1

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					other), type of surgery, follow-up duration, baseline back pain, leg pain, COMI score; II. General linear model (unadjusted)	(-0.19 to 0.39), SE = 0.15, p = .497; d = 0.03 (-0.06 to 0.12);
		NRS (0-10) leg pain mean reduction				I. 6-12 months M = 3.4, SD = 3.4, >12 months M = 3.5, SD = 3.6; *MD = 0.1 (-0.22 to 0.42), SE = 0.16, p = .538, d = 0.03 (-0.06 to 0.12); II. 6-12 months M = 3.4, SD = 3.4, >12 months M = 3.3, SD = 3.6; *MD = -0.1 (-0.42 to 0.22), SE = 0.16, p = .538, d = 0.03 (-0.06 to 0.12)
Støttrup 2019 [56]	Duration of leg pain (months): 3-12 vs >12	VAS (0-100) mean back pain reduction	12	1531 (1095 3-12 months, 436	I. Analysis of variance (unadjusted)	I. 3-12 months M = 21.32 (19.13 to 23.52), >12 months M = 17.23 (13.83 to 20.62); *MD = -4.09 (-8.18 to -

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
				>12 months)		0.002), SE = 2.08, p = .05; d = -0.11 (-0.22 to 0.00)
		VAS (0-100) mean leg pain reduction				I. 3-12 months M = 45.10 (42.87 to 47.34), >12 months M = 35.21 (31.51 to 38.91); *MD = -9.89 (-14.14 to -5.64), SE = 2.16, p < .001; d = -0.26 (-0.37 to -0.15)
Hegarty 2012 [9]	Duration of pain (months)	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53	I. Mann-Whitney U test (unadjusted)	I. No PPSP Mdn = 6, range 3 – 48, PPSP Mdn = 9, range 3 - 60; U = 282, z = 0.89, p = .36
	Pain intensity					
Anderson 2006 [53]	VAS (0-10) pain	VAS (0-10) 30% pain reduction	24 (M = 30)	106	I. Multivariate logistic regression; adjusted for: baseline work status (working), smoking (yes),	I. OR = 0.76 (0.58 to 1.00), p = .049

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					gender (NR), worker's compensation (yes), age <48, age >48, baseline pain (0-10 VAS), levels fused (single), cage type (BAK)	
Hegarty 2012 [9]	Present pain intensity (0-5; scale not clear)	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53	I. Multivariate logistic backward stepwise regression; adjusted for: age (years), disability (RMDQ); II. Mann-Whitney U test (unadjusted); III. Spearman's rho correlation (unadjusted)	I. b = .53, SE = .34, Wald = 2.6, p = .17, OR = 1.67, chi-square = 2.8 p = .09; II. No PPSP Mdn = 2, range = 0 – 5, PPSP Mdn = 2, range 1 - 5; U = 320, z = -1.8, p = .85, r = -.25; III. rho = .11, p = .43
	Preoperative pain severity (0-10 VAS)				I. Mann-Whitney U test (unadjusted); II. Spearman's rho correlation (unadjusted)	I. No PPSP Mdn = 6.5, range 1 – 10, PPSP Mdn = 5.7, range 0 - 7; U = 278, z = -0.95, p = .34, r = -.13; II. rho = -.006, p = .96
	Pain quality					

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Hegarty 2012 [9]	Short form McGill Pain Questionnaire score	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53	I. Mann-Whitney U test (unadjusted); II. Spearman's rho correlation (unadjusted)	I. No PPSP Mdn = 14, range 2 – 44, PPSP Mdn = 17, range 4 – 43; U = 237, z = -1.6, p = .11, r = -.24; II. rho = -.16, p = .26
	Night-time pain					
Gepstein 2007 [48]	Night-time pain vs no night-time pain	VAS (0-10) pain change	M = 46	220 (69 Arab, 151 Jewish)	I. Wilcoxon rank sum test (unadjusted)	I. Arab: r = -.43, p < .001; Jewish: r = -.38, p < .001
	Sensory detection threshold					
Hegarty 2012 [9]	Electrical sensory (mA) detection threshold on contralateral forearm and affected	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the	3	53	I. Mann-Whitney U test (unadjusted)	I. No PPSP Mdn = 8.7, range 0.9 - 17.2, PPSP Mdn = 6.9, range 1.2 - 25.3; U = 277, z = 0.9, p = .33

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	dermatome on the affected and contralateral lower limbs	past 2 weeks) vs PPSP (<70% reduction)				
	Pain detection threshold					
Hegarty 2012 [9]	Electrical pain (mA) detection threshold on contralateral forearm and affected dermatome on the affected and contralateral lower limbs	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53	I. Mann-Whitney U test (unadjusted)	I. No PPSP Mdn = 18.9, range 3.8 - 62.1, PPSP Mdn = 20.5, range 1.5 - 82.0; U = 283, z = 0.8, p = .39
Muller 2019 [58]	Electrical pain (mA) detection threshold on (a) single, (b) repeated stimulation	No FBSS (≥30% reduction in max. 0-10 NRS pain intensity during the last 7 days)	12	113	I. Multiple logistic regression; adjusted for: electrical pain detection on single (b) and repeated (a) stimulation (mA), pressure pain detection and tolerance on 2nd toe, 2nd finger, and most pain back	I. (a) single: OR = 1.54 (0.54 to 4.35), p = .42; (b) repeated: OR = 1.75 (0.65 to 4.55), p = .27

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
		vs FBSS (<30% reduction)			site (kPa), heat pain detection on leg and most pain back site (<50.5 °C), and cold pain detection thresholds on leg and most pain back site (>0.0 °C), cold pressor (hand withdrawal <120s), CPM (% without), type of surgery (instrumented / not), operated segments (multiple / single), gender (female), catastrophizing (PCS), BMI, Lasègue sign (positive), finger-floor distance (>10cm), baseline disability (ODI), non-opioid (yes) and opioid (yes) analgesics	
	Pressure pain (kPa) detection threshold on (a) 2nd toe, (b) 2nd finger, (c) most painful back site				I. Multiple logistic regression; adjusted for: electrical pain detection on single and repeated stimulation (mA), pressure pain detection on 2nd toe (b, c), 2nd finger (a, c), and most pain back site (a, b) and tolerance (kPa), heat pain detection on leg and most pain back site (<50.5 °C), and cold pain detection thresholds on leg and most pain	I. (a) 2nd toe: OR = 0.50 (0.18 to 1.45), p = .20; (b) 2nd finger: OR = 0.59 (0.24 to 1.47), p = .26; (c) back: OR = 1.00 (0.37 to 2.78), p = 1.00

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					back site (>0.0 °C), cold pressor (hand withdrawal <120s), CPM (% without), type of surgery (instrumented / not), operated segments (multiple / single), gender (female), catastrophizing (PCS), BMI, Lasègue sign (positive), finger-floor distance (>10cm), baseline disability (ODI), non-opioid (yes) and opioid (yes) analgesics	
	Heat pain detection threshold <50.5 °C vs ≥50.5 °C on (a) leg, (b) most painful back site				I. Multiple logistic regression; adjusted for: electrical pain detection on single and repeated stimulation (mA), pressure pain detection and tolerance on 2nd toe, 2nd finger, and most pain back site (kPa), heat pain detection on leg (b) and most pain back site (a) (<50.5 °C), and cold pain detection thresholds on leg and most pain back site (>0.0 °C), cold pressor (hand withdrawal <120s), CPM (% without), type of surgery (instrumented / not), operated segments (multiple	I. (a) leg: OR = 1.22 (0.41 to 3.57), p = .72; (b) back: OR = 1.67 (0.36 to 7.69), p = .51

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					/ single), gender (female), catastrophizing (PCS), BMI, Lasègue sign (positive), finger-floor distance (>10cm), baseline disability (ODI), non-opioid (yes) and opioid (yes) analgesics	
	Cold pain detection threshold >0.0 °C vs ≤0.0 °C on (a) leg, (b) most painful back site				I. Multiple logistic regression; adjusted for: electrical pain detection on single and repeated stimulation (mA), pressure pain detection and tolerance on 2nd toe, 2nd finger, and most pain back site (kPa), heat pain detection on leg and most pain back site (<50.5 °C), and cold pain detection thresholds on leg (b) and most pain back site (a) (>0.0 °C), cold pressor (hand withdrawal <120s), CPM (% without), type of surgery (instrumented / not), operated segments (multiple / single), gender (female), catastrophizing (PCS), BMI, Lasègue sign (positive), finger-floor distance (>10cm), baseline	I. (a) leg: OR = 0.71 (0.26 to 1.96), p = .51; (b) back: OR = 0.78 (0.29 to 2.08), p = .62

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					disability (ODI), non-opioid (yes) and opioid (yes) analgesics	
	Pain tolerance threshold					
Muller 2019 [58]	Pressure pain (kPa) tolerance threshold on (a) 2nd toe, (b) 2nd finger, (c) most painful back site	No FBSS (≥30% reduction in max. 0-10 NRS pain intensity during the last 7 days) vs FBSS (<30% reduction)	12	113	I. Multiple logistic regression; adjusted for: electrical pain detection on single and repeated stimulation (mA), pressure pain tolerance on 2nd toe (b, c), 2nd finger (a, c), and most pain back site (a, b) and detection (kPa), heat pain detection on leg and most pain back site (<50.5 °C), and cold pain detection thresholds on leg and most pain back site (>0.0 °C), cold pressor (hand withdrawal <120s), CPM (% without), type of surgery (instrumented / not), operated segments (multiple / single), gender (female), catastrophizing (PCS), BMI, Lasègue sign (positive), finger-floor distance (>10cm), baseline disability (ODI), non-opioid (yes) and opioid (yes) analgesics	I. (a) 2nd toe: OR = 1.00 (0.38 to 2.63), p = 1.00; (b) 2nd finger: OR = 1.27 (0.43 to 3.70), p = .66; (c) back: OR = 0.86 (0.30 to 2.44), p = .78

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	Cold pressor test (1.5 °C), hand withdrawal time <120sec vs ≥120sec				I. Multiple logistic regression; adjusted for: electrical pain detection on single and repeated stimulation (mA), pressure pain detection and tolerance on 2nd toe, 2nd finger, and most pain back site (kPa), heat pain detection on leg and most pain back site (<50.5 °C), and cold pain detection thresholds on leg and most pain back site (>0.0 °C), CPM (% without), type of surgery (instrumented / not), operated segments (multiple / single), gender (female), catastrophizing (PCS), BMI, Lasègue sign (positive), finger-floor distance (>10cm), baseline disability (ODI), non-opioid (yes) and opioid (yes) analgesics	I. OR = 1.20 (0.32 to 4.55), p = .78
Hegarty 2012 [9]	Electrical pain (mA) tolerance threshold on contralateral forearm and affected	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min	3	53	I. Mann-Whitney U test (unadjusted)	I. No PPSP Mdn = 26.6, range 7 - 95.8, PPSP Mdn = 34.1, range 3.1 - 99; U = 277, z = 0.9, p = .34

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	dermatome on the affected and contralateral lower limbs	walking test] in the past 2 weeks) vs PPSP (<70% reduction)				
	Conditioned pain modulation (CPM)					
Muller 2019 [58]	CPM, percentage of participants without increase of pressure pain detection threshold on 2nd toe (test stimulus) after cold pressor test (conditioning stimulus)	No FBSS (≥30% reduction in max. 0-10 NRS pain intensity during the last 7 days) vs FBSS (<30% reduction)	12	113	I. Multiple logistic regression; adjusted for: electrical pain detection on single and repeated stimulation (mA), pressure pain detection and tolerance on 2nd toe, 2nd finger, and most pain back site (kPa), heat pain detection on leg and most pain back site (<50.5 °C), and cold pain detection thresholds on leg and most pain back site (>0.0 °C), cold pressor (hand withdrawal <120s), type of surgery (instrumented / not), operated segments (multiple / single), gender (female), catastrophizing (PCS), BMI, Lasègue sign (positive), finger-floor	I. OR = 0.85 (0.19 to 3.70), p = .83

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					distance (>10cm), baseline disability (ODI), non-opioid (yes) and opioid (yes) analgesics	
	Disability / physical function					
Hegarty 2012 [9]	Physical disability due to low back pain (RMDQ score; 0-24)	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53	I. Multivariate logistic backward stepwise regression; adjusted for: age (years), present pain intensity	I. b = -.22, SE = .11, Wald = 4.5, p = .03, OR = 0.83, chi-square = 5.9, p = .015;
					II. Mann-Whitney U test (unadjusted)	II. No PPSP Mdn = 16.5, rage 3 – 23, PPSP Mdn = 17.5, range 8 - 23; U = 230, z = -1.8, p = .06, r = -.25
	Physical Component Score (PCS; SF-36)				I. Independent t-test (unadjusted) II. Spearman's rho correlation (unadjusted)	I. No PPSP M = 30.9, SE = 6.4, PPSP M = 32.9, SE = 5.9, p = .1; *d = -0.06 (-0.62 to 0.50); II. rho = -.24, p = .08
Anderson 2006 [53]	RMDQ score (0-24)	VAS (0-10) 30% pain reduction	24 (M = 30)	106	I. Multivariate logistic regression; adjusted for: baseline work status (working), gender (NR), worker's compensation (yes), age <48, age >48,	I. OR = 1.06 (0.95 to 1.20), p = .30

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					baseline pain (0-10 VAS), baseline disability (RMDQ), levels fused (single), cage type (BAK)	
Patel 2019 [51]	PROMIS PF (0-100) disability subgroup: mild (40-50) vs moderate (30-39.9) vs severe (20-29.9)	VAS (0-100) back pain change VAS (0-100) leg pain change	12	130 (20 mild, 83 moderate, 27 severe)	I. Multiple linear regression; adjusted for BMI (obesity ≥30 kg/m^2), worker's compensation insurance (yes)	I. Mild M = -4.1, moderate M = -3.7, severe M = -2.2; p = .222 I. Mild M = -2.8, moderate M = -3.1, severe M = -3.1; p = .229
Sigmundsson 2012 [52]	Walking distance >1000m vs 500-1000m vs 100-499m vs <100m	VAS (0-100) leg pain reduction	12	76	I. Multivariable regression; adjusted for: age, baseline leg and back pain (0-100 VAS), duration of leg and back pain (>24 months), spondylolisthesis (yes), stenosis (multilevel)	I. >1000m walking distance: B = -72 (-107 to -37), p < .001
	Physical comorbidity					
Gepstein 2007 [48]	Number of comorbidities	VAS (0-10) pain change	M = 46	220 (69 Arab, 151 Jewish)	I. Multiple linear regression; adjusted for: [Jewish] diabetes (yes), osteoarthritis (yes), total joint replacement (yes), education (years); II. Spearman's rank correlation (unadjusted)	I. Jewish: b = -0.31, SE = 0.13, beta = -0.19, t = -2.46, p = .015 (R^2 = 0.44); II. Arab: r = -.24, p < .05; Jewish: r = -.25, p < .01

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	Having vs not having peripheral arterial disease (ICD-9 code)				I. Multiple linear regression; adjusted for: [Arab] age, diabetes (yes), osteoarthritis (yes);	I. Arab: b = -1.26, SE = 0.55, beta = -0.20, t = -2.28, p = .026 (R^2 = 0.27);
					II. Wilcoxon rank sum test (unadjusted)	II. Arab: r = -.41, p < .001; Jewish: r = -.35, p < .001
	Having vs not having diabetes (ICD-9 code)				I. Multiple linear regression; adjusted for: [Arab] age, osteoarthritis (yes), peripheral arterial disease (yes) / [Jewish] osteoarthritis (yes), total joint replacement (yes), education (years), comorbidities (number);	I. Arab: b = -2.07, SE = 0.46, beta = -0.42, t = -4.51, p < .001 (R^2 = 0.27); Jewish: b = -2.11, SE = 0.54, beta = -0.31, t = -3.93, p < .001 (R^2 = 0.44);
					II. Wilcoxon rank sum test (unadjusted)	II. Arab: r = -.40, p < .001; Jewish: r = -.49, p < .001
	Having vs not having osteoarthritis (ICD-9 code)				I. Multiple linear regression; adjusted for: [Arab] age, diabetes (yes), peripheral arterial disease (yes) / [Jewish] diabetes (yes), total joint replacement (yes), education (years), comorbidities (number);	I. Arab: b = -1.34, SE = 0.54, beta = -0.23, t = -2.48, p = .016 (R^2 = 0.27); Jewish: b = -2.29, SE = 0.61, beta = -0.32, t = -3.75, p < .001 (R^2 = 0.44);

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					II. Wilcoxon rank sum test (unadjusted)	II. Arab: r = -.27, p < .05; Jewish: r = -.55, p < .001
	Having vs not having total joint replacement (ICD-9 code)				I. Multiple linear regression; adjusted for: [Jewish] diabetes (yes), osteoarthritis (yes), education (years), comorbidities (number);	I. Jewish: b = -2.46, SE = 0.85, beta = -0.23, t = -2.89, p = .004 (R^2 = 0.44);
					II. Wilcoxon rank sum test (unadjusted)	II. Arab: r = .06, p > .05; Jewish: r = .51, p < .001
	ASA class				I. Wilcoxon rank sum test (unadjusted)	II. Arab: r = -.03, p > .05; Jewish: r = -.29, p < .001
	BMI					
Gepstein 2007 [48]	BMI (kg/m^2)	VAS (0-10) pain change	M = 46	220 (69 Arab, 151 Jewish)	I. Spearman's rank correlation (unadjusted)	I. Arab: r = -.25, p < .05; Jewish: r = -.30, p < .001
	Smoking					
Anderson 2006 [53]	Smoking vs not smoking	VAS (0-10) 30% pain reduction	24 (M = 30)	106	I. Multivariate logistic regression; adjusted for: baseline work status (working), smoking (yes), gender (NR), worker's compensation (yes), age	I. OR = 1.14 (0.42 to 3.13), p = .80

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					<48, age >48, baseline disability (RMDQ), levels fused (single), cage type (BAK)	
	Psychological					
	Pain catastrophizing					
Kim 2015b [50]	PCS total score (0-50): low <25 vs high ≥25	VAS (0-10) leg pain reduction VAS (0-10) back pain reduction	3, 6, 12	138 (68 low PCS, 70 high PCS)	I. Mixed model for repeated measures (unadjusted): Fixed effects: PCS (low / high), Time (baseline / 3 / 6 / 12 months), PCS x Time interaction; Random effect: Subject	I. PCS group p = .040; Time p < .001; PCS x Time p = .820 I. PCS group p < .001; Time p < .001; PCS x Time p = .030
Hegarty 2012 [9]	PCS a) total score; b) helplessness subscale; c) rumination subscale; d) magnification subscale	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53	I. Independent t-test (unadjusted); II. Spearman's rho correlation (unadjusted)	a) I. No PPSP M = 31.6, SE = 14.9, PPSP M = 43.9, SE = 12.6; t(51) = -3.0, p = .004, r = -.38; II. rho = -.26, p = .06; b) I. No PPSP M = 14.3, SE = 7.2, PPSP M = 20, SE = 6.6; p < .05; *d = -0.15 (-0.71 to 0.40) c) I. No PPSP M = 5.6, SE = 3.9, PPSP M = 9.3, SE = 3.2; p < .001; *d = 0.55 (-

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
						0.74 to 0.37) d) I. No PPSP M = 10.8, SE = 5.2, PPSP M = 15.3, SE = 4.4; p < .05; *d = -0.17 (-0.73 to 0.39)
	Pain sensitivity					
Kim 2015a [49]	PSQ total score (0-10) low <6.5 vs high ≥ 6.5	VAS (0-10) leg pain reduction	3, 6, 12	171 (87 low PSQ, 84 high PSQ)	I. Mixed model for repeated measures (unadjusted): Fixed effects: PSQ (low/high), Time (baseline / 3 / 6 / 12 months), PSQ x Time interaction; Random effect: Subject	I. PSQ group: p < .001; Time p < .001 *(no contrasts)*; PSQ x time interaction p = .950
		VAS (0-10) back pain reduction	3, 6, 12	171 (87 low PSQ, 84 high PSQ)		I. PSQ group p < .001 (overall); Time p < .001 *(no contrasts)*; PSQ x time interaction p = .126
		VAS (0-10) leg pain mean percent reduction	12	124 (64 low PSQ, 60 high PSQ)	I. Independent t-test (unadjusted)	I. 51.90% (Low PSQ), 38.87% (High PSQ) decrease, p = .206.
		VAS (0-10) back pain mean percent reduction	12	124 (64 low PSQ, 60 high PSQ)		I. 42.76% (Low PSQ), 34.55% (High PSQ) decrease, p = .398

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	Pain Drawing					
Ekman 2009 [55]	Pain Drawing: organic (organic, possibly organic) vs non-organic (non-organic, possibly non-organic)	Pain Index (0-100) mean reduction	24	164 (126 organic, 38 non-organic)	I. Mann-Whitney U test (unadjusted)	I. Organic M = 31, non-organic M = 23, p = .09
	Mental functioning					
Hegarty 2012 [9]	Mental Component Score (MCS; SF-36)	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53	I. Independent t-test (unadjusted)	I. No PPSP M = 43.7, SE = 11.6, PPSP M = 37.4, SE = 9.9; p = .1, *d = 0.11 (-0.45 to 0.66)
	Anxiety					

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Hegarty 2012 [9]	Anxiety subscale score (HADS)	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53	I. Independent t-test (unadjusted); II. Spearman's rho correlation (unadjusted)	I. No PPSP M = 6.2, SE = 2.9, PPSP M = 8.5, SE = 3.9; t(51) = -2.4, p = .02, r = -.31; II. rho = -.18, p = .19
	Depression					
Hegarty 2012 [9]	Depression subscale score (HADS)	No PPSP (≥70% reduction in 0-10 VAS movement related pain intensity [5min walking test] in the past 2 weeks) vs PPSP (<70% reduction)	3	53	I. Independent t-test (unadjusted); II. Spearman's rho correlation (unadjusted)	I. No PPSP M = 6.6, SE = 3.9, PPSP M = 8.5, SE = 4.6; t(51) = -1.47, p = .14, r = -.1; II. rho = -.17, p = .22

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition*[a]	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors*[b] */ unadjusted*	*Estimate (95% CI)*[c]
Gepstein 2007 [48]	Depression vs no depression (ICD-9 code)	VAS (0-10) pain change	M = 46	220 (69 Arab, 151 Jewish)	I. Wilcoxon rank sum test (unadjusted)	I. Arab: r = -.30, p < .05; Jewish: r = -.32, p < .001

ASA, American Society of Anesthesiologists; BAK, Bagby and Kuslich cage; BMI, body mass index; COMI, Core Outcome Measures Index; CPM, conditioned pain modulation; d, Cohen's d (standardized mean difference); FBSS, failed back surgery syndrome; HADS, Hospital Anxiety and Depression Scale; HNP, herniated nucleus pulposus; ICD-9, International Classification of Diseases, 9th revision; L1-5, lumbar spine segment; M, mean; MD, mean difference; Mdn, median; NR, not reported; NRS, Numeric Rating Scale; ODI, Oswestry Disability Index; OR, odds ratio; PCS, Pain Catastrophizing Scale; PPSP, persistent post-surgical pain; PROMIS PF, Patient-Reported Outcomes Measurement Information System - Physical Function; PSQ, Pain Sensitivity Questionnaire; RMDQ, Roland Morris Disability Questionnaire; S1, sacral spine segment; SD, standard deviation; SE, standard error; SF-36 PCS / MCS, Short Form Health Survey - Physical / Mental Component Summary; VAS, Visual Analog Scale.

[a] Greater disability indicated by higher ODI, RMDQ, and COMI scores, and lower SF-36, and PROMIS scores; greater pain intensity indicated by higher VAS and NRS scores. Reduction / improvement in outcome is from preoperative baseline to follow-up. For consistency, outcome definitions (and direction of effect estimates) were reversed in studies reporting prediction of a *failure* rather than success of surgery (Hegarty 2012, Muller 2019).

[b] If categories not specified, factor analyzed as a continuous variable.

[c] Statistics and 95% confidence intervals reported where available. * Effects estimated from available data where possible. [fig], results extracted from a figure.

Table 4. Effects of prognostic factors on change in disability from baseline to the last available follow-up.

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	Sociodemographic					
	Age					
Anderson 2006 [53]	Age <48 years vs >48 years	RMDQ (0-24) 30% reduction	24 (M = 30)	93	I. Multivariate logistic regression; adjusted for: baseline work status (working), smoking (yes), gender (?), worker's compensation (yes), age >48, baseline pain (0-10 VAS), baseline disability (RMDQ), levels fused (single), cage type (BAK)	I. Age <48 years: OR = 0.88 (0.30 to 2.61), p = .82; Age >48 years: OR = 1.28 (0.32 to 5.07), p = .72
	Gender					
Anderson 2006 [53]	NR (reference level not specified – direction of the effect uncertain)	RMDQ (0-24) 30% reduction	24 (M = 30)	93	I. Multivariate logistic regression; adjusted for: baseline work status (working), smoking (yes), worker's compensation (yes), age <48, age >48, baseline pain (0-10 VAS), baseline disability (RMDQ), levels fused (single), cage type (BAK)	I. OR = 1.27 (0.49 to 3.31), p = .62
Stromqvist 2008 [57]	Female vs male	SF-36 physical function subscale (0-	12	301 (136 female, 165 male)	I. Analysis of covariance (unadjusted – covariates not specified)	I. Female M = 27, male M = 31, p > .05

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
		100) mean improvement				
Watkins 1986 [65]	Male vs female	ODI (0-100), ≥17 vs <17 reduction	>24	42 (23 male, 19 female)	I. 2 x 2 frequency table (unadjusted)	I. 14 male, 8 female ≥17 reduction; *OR = 2.14 (0.62 to 7.37)
	Ethnicity					
Gepstein 2007 [48]	Israeli Arabs vs Israeli Jews	Barthel index (0-100) mean change	M = 46	220 (69 Arab, 151 Jewish)	I. Independent samples t-test (unadjusted)	I. Arab M = -9.86, SD = 11.01, Jewish M = -9.20, SD = 8.49, p > .05, *d = 0.07 (-0.21 to 0.36)
	Work status					
Anderson 2006 [53]	Working (including home working and studies) vs not working	RMDQ (0-24) 30% reduction	24 (M = 30)	93	I. Multivariate logistic regression; adjusted for: smoking (yes), gender (NR), worker's compensation (yes), age <48, age >48, baseline pain (0-10 VAS), baseline disability (RMDQ), levels fused (single), cage type (BAK)	I. OR = 0.62 (0.23 to 1.73), p = .36
	Worker's compensation					

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Anderson 2006 [53]	Compensation claim vs no compensation	RMDQ (0-24) 30% reduction	24 (M = 30)	93	I. Multivariate logistic regression; adjusted for: baseline work status (working), smoking (yes), gender (NR), age <48, age >48, baseline pain (0-10 VAS), baseline disability (RMDQ), levels fused (single), cage type (BAK)	I. OR = 1.61 (0.59 to 4.39), p = .35
	Health-related					
	Sciatica					
Ekman 2009 [55]	Sciatica (pain symbols below the knee on Pain Drawing) vs no sciatica	Disability Rating Index (0-100) mean reduction	24	164 (119 sciatica, 45 no sciatica)	I. Mann-Whitney U test (unadjusted)	I. Sciatica M = 17, no sciatica M = 18, p = .82
	Symptom duration					
Radcliff 2011 [61] (Spinal stenosis)	<12 months vs ≥12 months (at enrolment)	SF-36 Physical Functioning (PF; 0-100) improvement	48	413 (255 <12 months,	I. Mixed model for repeated measures: Fixed effects: Symptom duration (<12m / ≥12m); Time-varying covariate: Treatment (surgery / non-	I. <12 months M = 24.7, SE = 1.9, ≥12 months M = 16.9, SE = 1.8, p(MD) = .002, *d = -0.28 (-0.48 to -0.08)

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
		ODI (0-100; MODEMS version) reduction		158 ≥12 months)	surgical, time baseline / 12 / 24 / 36 / 48 months); Random effect: Subject; Adjusted for: age, gender (female), BMI, race (white), smoking status (smoker), compensation status (any), comorbidities (joint, stomach, bowel, intestinal problems, osteoporosis, other - present), number of moderately/severely stenotic levels (0 / 1 / 2 / 3+), self-assessed baseline health trend (getting better / staying the same / getting worse), treatment preference (non-surgical / not sure / surgery), baseline stenosis bothersomeness, baseline score on outcome measure, center (NR)	I. <12 months M = -22.3, SE = 1.5, ≥12 months M = -16.2, SE = 1.4, p(MD) = .002, *d = 0.28 (0.08 to 0.48)
Radcliff 2011 [61] (Degenerative	<12 months vs ≥12 months (at enrolment)	SF-36 Physical Functioning (PF; 0-100) improvement	48	391 (254 <12 months,	I. Mixed model for repeated measures (as above)	I. <12 months M = 26.6, SE = 1.9, ≥12 months M = 25.8, SE = 1.7, p(MD) = .74, *d = -0.03 (-0.24 to 0.18)

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
spondylolisthesis)		ODI (0-100; MODEMS version) reduction		137 ≥12 months)		I. <12 months M = -23.6, SE = 1.4, ≥12 months M = -22.1, SE = 1.3, p(MD) = .44, *d = 0.08 (-0.13 to 0.28)
Tampin 2020 [64]	Symptom duration (months)	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Independent t-test (unadjusted)	I. ≥30% M = 11.4, SD = 7.6, <30% M = 12.4, SD = 7.0, p = .751, *d = -0.13 (-0.94 to 0.67)
Støttrup 2019 [56]	Duration of leg pain (months): 3-12 vs >12	ODI (0-100) mean reduction	12	1531 (1095 3-12 months, 436 >12 months)	I. Analysis of variance (unadjusted)	I. 3-12 months M = 25.96 (24.49 to 27.43), >12 months M = 19.68 (17.41 to 21.94); *MD = -6.28 (-9.02 to -3.54), SE = 1.39, p < .001; d = -0.26 (-0.37 to -0.14)
Cushnie 2019 [54]	Duration of the main neurologic leg complaint	ODI (0-100) reduction	12	466 (26 3-6 months, 85 6-12 months, 125	I. Wilcoxon signed-rank test (unadjusted)	I. 3-6 months M = -24 (-16 to -33); 6-12 months M = -21 (-17 to -25); 12-24 months M = -17 (-14 to -21); >24 months M = -15 (-13 to -17) [fig]

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	(months): 3-6 vs 6-12 vs 12-24 vs >24	SF-12 Physical Component Score (PCS; 0-100) improvement		12-24 months, 230 >24 months)		*overlapping CIs - no significant differences* I. 3-6 months M = 12 (7 to 17); 6-12 months M = 10 (8 to 12.5); 12-24 months M = 9.5 (8 to 11.5); >24 months M = 8 (7 to 9) [fig] *overlapping CIs - no significant differences*
Zweig 2017 [10]	Duration of pre-operative conservative treatment (months): 6-12 vs >12	COMI (0-10) ≥2 vs <2 points change	3 – 30 (M = 17)	2016 (758 6-12 months, 1258 >12 months)	I. Chi-square test after propensity score weighing adjustment for: age, gender (female), ASA (1 / 2 / >2), number of affected segments (1 / 2–3 / >3), level operated (L1/2–L2/3 / L3/4 / L4/5 / L5/S1), surgical goal (pain reduction / functional / neurological improvement), patient-reported main problem (back / leg pain / sensory disturbances /	I. 497 6-12 months, 825 >12 months had ≥2 change; *OR = 1.00 (0.83 to 1.21); II. OR = 1.00 (0.83 to 1.20) [fig]; III. 497 6-12 months, 823 >12 months had ≥2 change; *OR = 1.01 (0.83 to 1.22)

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					other), type of surgery, follow-up duration, baseline back pain, leg pain, COMI score; II. Multiple logistic regression; adjusted for inverse probability of treatment weight (propensity score), sequestrectomy (yes), foraminotomy (yes); III. Chi-square test (unadjusted)	
		COMI (0-10) mean reduction			I. General linear model after propensity score weighing adjustment for: age, gender (female), ASA (1 / 2 / >2), number of affected segments (1 / 2–3 / >3), level operated (L1/2–L2/3 / L3/4 / L4/5 / L5/S1), surgical goal (pain reduction / functional / neurological improvement), patient-reported main problem (back / leg pain / sensory disturbances / other), type of surgery, follow-up duration, baseline back pain, leg pain, COMI score;	I. 6-12 months M = 3.4, SD = 2.8, >12 months M = 3.4, SD = 2.9; *MD = 0 (-0.26 to 0.26), SE = 0.13, p = 1.00, d = 0 (-0.09 to 0.09) II. 6-12 months M = 3.4, SD = 2.8, >12 months M = 3.3, SD = 2.9; *MD = -0.1 (-0.36 to 0.16), SE = 0.13, p = .448, d = -0.04 (-0.13 to 0.06)

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					II. General linear model (unadjusted)	
	Pain intensity					
Anderson 2006 [53]	VAS (0-10) pain	RMDQ (0-24) 30% reduction	24 (M = 30)	93	I. Multivariate logistic regression; adjusted for: baseline work status (working), smoking (yes), gender (NR), worker's compensation (yes), age <48, age >48, baseline pain (0-10 VAS), levels fused (single), cage type (BAK)	I. OR = 1.10 (0.85 to 1.41), p = .48
Tampin 2020 [64]	Average leg pain intensity (0-10 NRS) over last week	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Independent t-test (unadjusted)	I. ≥30% M = 5.9, SD = 1.8, <30% M = 4.3, SD = 2.3, p = .041, *d = 0.85 (0.04 to 1.67)
	Average leg pain intensity (0-10 NRS) over last 24 hours					I. ≥30% M = 6.0, SD = 2.0, <30% M = 4.4, SD = 1.9, p = .057, *d = 0.81 (-0.01 to 1.62)

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	Average back pain intensity (0-10 NRS) over last week					I. ≥30% M = 4.4, SD = 2.5, <30% M = 3.8, SD = 3.1, p = .599, *d = 0.23 (-0.57 to 1.04)
	Average back pain intensity (0-10 NRS) over last 24 hours					I. ≥30% M = 4.7, SD = 2.5, <30% M = 4.3, SD = 2.7, p = .700, *d = 0.16 (-0.64 to 0.96)
	Pain bothersomeness					
Tampin 2020 [64]	Leg pain bothersomeness (0-5) over last 2 weeks	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Independent t-test (unadjusted)	I. ≥30% M = 2.8, SD = 0.8, <30% M = 2.3, SD = 0.9, p = .122, *d = 0.61 (-0.20 to 1.43)
	Back pain bothersomeness (0-5) over last 2 weeks					I. ≥30% M = 2.1, SD = 1.0, <30% M = 1.8, SD = 0.7, p = .581, *d = 0.31 (-0.49 to 1.11)
	Neuropathic pain component					

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Tampin 2020 [64]	PainDETECT score (-1 – 38)	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Independent t-test (unadjusted)	I. ≥30% M = 15.8, SD = 5.4, <30% M = 15.3, SD = 4.4, p = .821, *d = 0.10 (-0.71 to 0.90)
	Sensory detection threshold					
Tampin 2020 [64]	Mechanical detection threshold on most painful back site (sensory loss z-score)	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Multiple logistic regression; adjusted for: gender (female), anxiety, depression, pain catastrophizing; II. Independent t-test (unadjusted)	I. OR = 2.63 (1.09 to 6.37), p = .032; II. ≥30% z = -1.8, SE = 0.3 (-2.2 to -1.6), <30% z = -0.6, SE = 0.25 (-0.9 to -0.4) [fig], p = .008
	Disability					
Anderson 2006 [53]	RMDQ score (0-24)	RMDQ (0-24) 30% reduction	24 (M = 30)	93	I. Multivariate logistic regression; adjusted for: baseline work status (working), gender (NR), worker's compensation (yes), age <48, age >48, baseline pain (0-10 VAS), baseline disability (RMDQ), levels fused (single), cage type (BAK)	I. OR = 1.02 (0.91 to 1.14), p = .78

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Schade 1999 [63]	RMDQ score (0-24)	RMDQ (0-24) reduction	24 (23-30)	42	I. Stepwise multiple regression analysis; adjusted for: extent of neural compromise (none / minor / major), job-related resignation	I. Beta = 0.33, T = 2.87, $p < .01$; final model: $F(3,39) = 13.1$, $p < .001$, $R = 0.71$, $R^2 = 0.50$, adj. $R^2 = 0.46$
Patel 2019 [51]	PROMIS PF (0-100) disability subgroup: mild (40-50) vs moderate (30-39.9) vs severe (20-29.9)	PROMIS Physical Function (0-100) score change	12	130 (20 mild, 83 moderate, 27 severe)	I. Multiple linear regression; adjusted for BMI (obesity ≥30 kg/m^2), worker's compensation insurance (yes)	I. Mild M = 11.1, moderate M = 10.0, severe M = 10.1; p = .012
		ODI (0-100) change				I. Mild M = -23.7, moderate M = -19.9, severe M = -17.0; p = .497
		SF-12 Physical Component Score (PCS; 0-100) change				I. Mild M = 15.4, moderate M = 10.1, severe M = 9.6; p = .040
Tampin 2020 [64]	ODI score (0-100)	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Independent t-test (unadjusted)	I. ≥30% M = 18.4, SD = 6.2, <30% M = 15.1, SD = 5.6, p = .201, *d = 0.54 (-0.27 to 1.35)

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	SF-36 Physical Component Score (0-100)				I. Mann-Whitney U test (unadjusted)	I. ≥30% Mdn = 35.6, IQR = 7.8, <30% Mdn = 38.4, IQR = 11.4, range 26-49, p = .104
	BMI					
Rihn 2012 [62] (Spinal stenosis)	Non-obese (BMI <30) vs obese (BMI≥30)	SF-36 PF (0-100) improvement	48	413 (250 non-obese, 163 obese)	I. Mixed model for repeated measures: Fixed effects: Obesity (non-obese / obese); Time-varying covariate: Treatment (surgery / non-surgical, time baseline / 12 / 24 / 36 / 48 months); Random effect: Subject; Adjusted for: age, gender (female), race (white), smoking status (smoker), compensation status (any), comorbidities (joint, stomach, bowel and intestinal problems, osteoporosis, other - present), number of moderately/severely stenotic levels (0 / 1 / 2 / 3+), self-assessed baseline health trend (getting better / staying the same / getting worse), treatment	I. Non-obese M = 22.5, SE = 1.7, obese M = 18.2, SE = 2.1, p(MD) = .10, *d = -0.16 (-0.36 to 0.04)

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					preference (non-surgical / not sure / surgery), baseline stenosis bothersomeness, baseline score on outcome measure, center (NR)	
		ODI (0-100; MODEMS version) reduction				I. Non-obese M = -20.1, SE = 1.3, obese M = -17.6, SE = 1.6, p(MD) = .22, *d = 0.12 (-0.08 to 0.32)
Rihn 2012 [62] (Degenerative spondylolisthesis)	Non-obese (BMI <30) vs obese (BMI≥30)	SF-36 PF (0-100) improvement	48	391 (235 non-obese, 156 obese)	I. Mixed model for repeated measures (as above)	I. Non-obese M = 27.9, SE = 1.6, obese M = 22.1, SE = 2, p(MD) = .022, *d = -0.24 (-0.44 to -0.03)
		ODI (0-100; MODEMS version) reduction				I. Non-obese M = -23.1, SE = 1.2, obese M = -21.7, SE = 1.5, p(MD) = .46, *d = 0.08 (-0.16 to 0.31)

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
McGuire 2014 [60] (Spinal stenosis)	Non-obese (BMI < 30) vs obese (30 ≤ BMI < 35) vs extremely obese (BMI ≥ 35)	SF-36 PF (0-100) improvement	48	413 (250 non-obese, 104 obese, 59 extremely obese)	I. Mixed model for repeated measures: Fixed effects: Obesity (non-obese / obese / extremely obese); Time-varying covariate: Treatment (surgery / non-surgical, time baseline / 12 / 24 / 36 / 48 months); Random effect: Subject; Adjusted for: age, gender (female), race (white), smoking status (smoker), compensation status (any), comorbidities (joint, stomach, bowel and intestinal problems, osteoporosis, other - present), number of moderately/severely stenotic levels (0 / 1 / 2 / 3+), self-assessed baseline health trend (getting better / staying the same / getting worse), treatment preference (non-surgical / not sure / surgery), baseline stenosis bothersomeness, baseline score on outcome measure, center (NR); Wald test	I. Non-obese M = 22.5, SE = 1.7, obese M = 18.4, SE = 2.6, extremely obese M = 17.9, SE = 3.4, p = .26

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
		ODI (0-100; MODEMS version) reduction				I. Non-obese M = -20.1, SE = 1.3, obese M = -17.6, SE = 2, extremely obese M = -17.3, SE = 2.7, p = .46
McGuire 2014 [60] (Degenerative spondylolisthesis)	Non-obese (BMI < 30) vs obese (30 ≤ BMI < 35) vs extremely obese (BMI ≥ 35)	SF-36 PF (0-100) improvement	48	391 (235 non-obese, 90 obese, 66 extremely obese)	I. Mixed model for repeated measures (as above); Wald test	I. Non-obese M = 27.9, SE = 1.6, obese M = 22.8, SE = 2.5, extremely obese M = 21.2, SE = 3, p = .069
		ODI (0-100; MODEMS version) reduction				I. Non-obese M = -23.2, SE = 1.3, obese M = -22, SE = 2, extremely obese M = -21.3, SE = 2.4, p = .75
	Smoking					
Anderson 2006 [53]	Smoking vs not smoking	RMDQ (0-24) 30% reduction	24 (M = 30)	93	I. Multivariate logistic regression; adjusted for: baseline work status (working), smoking (yes), gender (NR), worker's compensation (yes), age	I. OR = 0.56 (0.21 to 1.54), p = .26

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
					<48, age >48, baseline disability (RMDQ), levels fused (single), cage type (BAK)	
	Sleep quality					
Tampin 2020 [64]	Sleep quality over last week (VAS 0-10)	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Mann Whitney U test (unadjusted)	I. ≥30% Mdn = 5.9, IQR = 3.9, <30% Mdn = 4.5, IQR = 3.2, range 0-10, p = .320
	Psychological					
	Pain catastrophizing					
Kim 2015b [50]	PCS total score (0-50): low <25, high ≥25	ODI (0-100) reduction	3, 6, 12	138 (68 low PCS, 70 high PCS)	I. Mixed model: Fixed effects: PCS (low / high), Time (baseline / 3 / 6 / 12 months), PCS x Time interaction; Random effect: Subject (unadjusted)	I. PCS group: p < .001; Time: p < .001; PCS x Time: p = .016 *(no contrasts)*
			12	103 (54 low PCS, 49 high PCS)	I. Paired t-test (unadjusted)	I. Low PCS M = 14.7, SD = 22.4, high PCS M = 22.5, SD = 17.7; d = -0.38 (-0.77 to 0.001), p = .053

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Tampin 2020 [64]	PCS total score	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Mann-Whitney U test (unadjusted)	I. ≥30% Mdn = 20.5, IRQ = 16.5, range NR, <30% Mdn = 13.0, IQR = 22.0, range 1-39, p = .328
	Pain sensitivity					
Kim 2015a [49]	PSQ total score (0-10) low <6.5 & high ≥6.5	ODI (0-100) reduction	3, 6, 12	171 (87 low PSQ, 84 high PSQ)	I. Mixed model for repeated measures: Fixed effects: PSQ (low/high), Time (baseline / 3 / 6 / 12 months), PSQ x Time interaction; Random effect: Subject (unadjusted)	I. PSQ group p < .001 (overall), Time p < .001, PSQ x Time interaction p = .757 *(no contrasts)*
			12	124 (64 low PSQ, 60 high PSQ)	I. Paired t-test (unadjusted)	I. Low PSQ M = 12.4, SD = 10.1, high PSQ M = 9.9, SD = 8.2; *d = -0.27 (-0.63 to 0.08)
	Pain Drawing					
Ekman 2009 [55]	Pain Drawing: organic vs non-organic	Disability Rating Index (0-100) mean reduction	24	164 (126 organic, 38 non-organic)	I. Mann-Whitney U test (unadjusted)	I. Organic M = 19, non-organic M = 11, p = .050

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
	Fear of movement					
Tampin 2020 [64]	Tampa Scale for Kinesiophobia	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Independent t-test (unadjusted)	I. ≥30% M = 44.5, SD = 6.6, <30% M = 42.6, SD = 9.4, p = .505, *d = 0.27 (-0.53 to 1.07)
	Mental functioning					
Tampin 2020 [64]	SF-36 Mental Component Score (0-100)	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Mann-Whitney U test (unadjusted)	I. ≥30% Mdn = 43.2, IQR = 14.7, range NR, <30% Mdn = 50.6, IQR = 16.3, range 28-62, p = .370
	Anxiety					
Tampin 2020 [64]	Anxiety subscale score (HADS)	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Mann-Whitney U test (unadjusted)	I. ≥30% Mdn = 8.0, IRQ = 5.2, range NR, <30% Mdn = 7.0, IQR = 7.0, range 2-16, p = .932
	Depression					
Tampin 2020 [64]	Depression subscale score (HADS)	ODI (0-100) ≥30% vs <30% reduction	12	48	I. Mann-Whitney U test (unadjusted)	I. ≥30% Mdn = 6.0, IRQ = 5.2, range NR, <30% Mdn = 7.0, IQR = 6.0, range 2-16, p = .775

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Schade 1999 [63]	Depression subscale from Psychological general well-being index	RMDQ (0-24) reduction	24 (23-30)	42	I. Parametric univariate analysis (unadjusted)	I. $p < .05$
	Vitality					
Schade 1999 [63]	Vitality subscale from Psychological general well-being index	RMDQ (0-24) reduction	24 (23-30)	42	I. Parametric univariate analysis (unadjusted)	I. $p < .05$
	Job-related resignation					
Schade 1999 [63]	4-item Job-related resignation scale (1-5 Likert ratings)	RMDQ (0-24) reduction	24 (23-30)	42	I. Stepwise multiple regression analysis; adjusted for: baseline disability (RMDQ), extent of neural compromise (none / minor / major);	I. Beta = 0.40, T = 3.53, $p < .001$; final model: $F(3,39) = 13.1$, $p < .001$, $R = 0.71$, $R^2 = 0.50$, adj. $R^2 = 0.46$;
					II. Parametric univariate analysis (unadjusted)	II. $p < .05$
	Neuroticism					

Study ID	Prognostic factor	Outcome		Sample size	Analysis	Effect estimates
	Measure, definition[a]	*Measure, definition[a]*	*Follow-up (months)*	*N analyzed*	*Method, adjusted for factors[b] / unadjusted*	*Estimate (95% CI)[c]*
Hagg 2003 [59]	Neuroticism subscale from Karolinska Scales of Personality standardized T score	ODI (0-100) ≥50% vs <50% reduction	24	183	I. Stepwise forward multiple regression analysis (unadjusted - single factor selected based on univariate analysis)	I. Beta = -0.096; OR = 0.91 (0.87 to 0.95); constant = 3.808

ASA, American Society of Anesthesiologists; BAK, Bagby and Kuslich cage; BMI, body mass index; COMI, Core Outcome Measures Index; d, Cohen's d (standardized mean difference); DRI, Disability Rating Index; HNP, herniated nucleus pulposus; L1-5, lumbar spine segment; M, mean; MD, mean difference; MODEMS, Musculoskeletal outcomes Data Evaluation and Management Systems; NR, not reported; ODI, Oswestry Disability Index; OR, odds ratio; PCS, Pain Catastrophizing Scale; PROMIS PF, Patient-Reported Outcomes Measurement Information System - Physical Function; PSQ, Pain Sensitivity Questionnaire; RMDQ, Roland Morris Disability Questionnaire; S1, sacral spine segment; SD, standard deviation; SE, standard error; SF-12/36 PF, Short Form Health Survey - Physical Functioning; VAS, Visual Analog Scale;

[a] Greater disability indicated by higher ODI, RMDQ, and COMI scores, and lower SF-36, and PROMIS scores; greater pain intensity indicated by higher VAS scores. Reduction / improvement in outcome is from pre-operative baseline to follow-up. For consistency, outcome definitions (and direction of effect estimates) were reversed in studies reporting prediction of a *failure* rather than success of surgery (Tampin 2020).

[b] If categories not specified, factor analyzed as a continuous variable.

[c] Statistics and 95% confidence intervals reported where available. * Effects estimated from available data where possible. [fig], results extracted from a figure.

Predictors of pain and disability outcomes following spinal surgery for chronic low back and radicular pain: A systematic review

Supplemental Digital Content

Contents

Supplemental Digital Content 1: Search strategies for electronic databases

A1. MEDLINE Ovid

1. exp Spine/su [Surgery]
2. ((spine or spinal) adj5 (surger* or surgical*)).ti,ab,kw.
3. 1 or 2
4. exp Lumbar Vertebrae/
5. exp Low Back Pain/
6. exp Lumbosacral Region/
7. ("low* back" or lumbar or lumbosacral).ti,ab,kw.
8. 4 or 5 or 6 or 7
9. 3 and 8
10. exp Lumbosacral Region/su [Surgery]
11. exp Lumbar Vertebrae/su [Surgery]
12. exp Low Back Pain/su [Surgery]
13. exp Sciatica/su [Surgery]
14. ((lumb* or "lower back") adj5 (surger* or surgical*)).ti,ab,kw.
15. 10 or 11 or 12 or 13 or 14
16. 9 or 15
17. exp chronic pain/
18. ((chronic or constant or persist* or longterm or long-term or "long standing" or longstanding or "long lasting" or longlasting) adj5 (pain or lbp)).ti,ab,kw.
19. (clbp or lumbar radiculopathy or lumbar radicular pain or sciatica or "postoperative pain").ti,ab,kw.
20. exp sciatica/
21. 17 or 18 or 19 or 20
22. 16 and 21
23. exp pain/
24. exp Recovery of Function/
25. pain*.ti,ab,kw.
26. "Visual Analogue Scale".ti,ab,kw.
27. VAS.ti,ab,kw.
28. "Numeric Rating Scale".ti,ab,kw.
29. "Numerical Rating Scale".ti,ab,kw.
30. NRS.ti,ab,kw.
31. "Numeric Pain Rating Scale".ti,ab,kw.
32. "Numerical Pain Rating Scale".ti,ab,kw.
33. NPRS.ti,ab,kw.
34. function*.ti,ab,kw.
35. disabilit*.ti,ab,kw.
36. ODI.ti,ab,kw.
37. "Oswestry Disability Index".ti,ab,kw.
38. "Oswestry Disability Questionnaire".ti,ab,kw.
39. "Roland and Morris Disability Index".ti,ab,kw.
40. "Roland and Morris Disability Questionnaire".ti,ab,kw.
41. RMDQ.ti,ab,kw.
42. RDQ.ti,ab,kw.
43. "Core Outcome Measures Index".ti,ab,kw.
44. COMI.ti,ab,kw.
45. exp return to work/
46. "return to work".ti,ab,kw.
47. "back to work".ti,ab,kw.
48. "work engagement".ti,ab,kw.
49. employment.ti,ab,kw.
50. or/23-49
51. 22 and 50
52. exp Predictive Value of Tests/

53. exp risk factor/
54. Predict*.ti,kw.
55. prognos*.ti,kw.
56. (Validat* or Rule*).mp.
57. (Predict* and (Outcome* or Risk* or Model* or value)).mp.
58. ((History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor*) and (Predict* or Model* or Decision* or Identif* OR Prognos*)).mp.
59. (Decision* and (Model* or Clinical*)).mp.
60. (Prognostic and (History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor* or Model* or value)).mp.
61. Logistic Models/
62. exp clinical decision rules/
63. Stratification.mp.
64. exp ROC Curve/
65. Discrimination.mp.
66. discriminate*.mp.
67. c-statistic.mp.
68. "c statistic".mp.
69. "Area under the curve".mp.
70. AUC.mp.
71. Calibration.mp.
72. Indices.mp.
73. Algorithm.mp.
74. Multivariable*.mp.
75. or/52-74
76. 51 and 75
77. animals/
78. humans/
79. 77 not 78
80. 76 not 79
81. case reports.pt.
82. 80 not 81
83. limit 82 to english language
84. limit 83 to yr="1984 -Current"

Notes

1. Lines 52-75 reflect an adapted Ingui filter (Ingui et al., 2001; Geersing et al., 2012), adapted to this specific study and Ovid format, and updated to include current indexing terms.

A2. Embase

1. exp spine surgery/
2. ((spine or spinal) adj5 (surger* or surgical*)).ti,ab,kw.
3. 1 or 2
4. exp lumbar vertebra/
5. exp low back pain/
6. exp lumbosacral region/
7. ("low* back" or lumbar or lumbosacral).ti,ab,kw.
8. 4 or 5 or 6 or 7
9. 3 and 8
10. exp lumbar vertebra/su [Surgery]
11. exp low back pain/su [Surgery]
12. exp sciatica/su [Surgery]
13. ((lumb* or "lower back") adj5 (surger* or surgical*)).ti,ab,kw.
14. 10 or 11 or 12 or 13
15. 9 or 14

16. exp chronic pain/
17. ((chronic or constant or persist* or longterm or long-term or "long standing" or longstanding or "long lasting" or longlasting) adj5 (pain or lbp)).ti,ab,kw.
18. (clbp or lumbar radiculopathy or lumbar radicular pain or sciatica or "postoperative pain").ti,ab,kw.
19. exp sciatica/
20. 16 or 17 or 18 or 19
21. 15 and 20
22. exp pain/
23. pain*.ti,ab,kw.
24. "Visual Analogue Scale".ti,ab,kw.
25. exp visual analog scale/
26. VAS.ti,ab,kw.
27. "Numeric Rating Scale".ti,ab,kw.
28. exp numeric rating scale/
29. "Numerical Rating Scale".ti,ab,kw.
30. NRS.ti,ab,kw.
31. "Numeric Pain Rating Scale".ti,ab,kw.
32. "Numerical Pain Rating Scale".ti,ab,kw.
33. NPRS.ti,ab,kw.
34. function*.ti,ab,kw.
35. disabilit*.ti,ab,kw.
36. ODI.ti,ab,kw.
37. "Oswestry Disability Index".ti,ab,kw.
38. "Oswestry Disability Questionnaire".ti,ab,kw.
39. "Roland and Morris Disability Index".ti,ab,kw.
40. "Roland and Morris Disability Questionnaire".ti,ab,kw.
41. exp Oswestry Disability Index/
42. RMDQ.ti,ab,kw.
43. RDQ.ti,ab,kw.
44. "Core Outcome Measures Index".ti,ab,kw.
45. COMI.ti,ab,kw.
46. exp return to work/
47. exp work resumption/
48. (return adj to adj work).ti,ab,kw.
49. (back adj to adj work).ti,ab,kw.
50. "work engagement".ti,ab,kw.
51. employment.ti,ab,kw.
52. or/22-51
53. 21 and 52
54. exp predictive value/
55. exp risk factor/
56. Predict*.ti,kw.
57. prognos*.ti,kw.
58. (Validat* or Rule*).ti,ab,kw.
59. (Predict* and (Outcome* or Risk* or Model* or value)).ti,ab,kw.
60. ((History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor*) and (Predict* or Model* or Decision* or Identif* or Prognos*)).ti,ab,kw.
61. (Decision* and (Model* or Clinical*)).ti,ab,kw.
62. (Prognostic and (History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor* or Model* or value)).ti,ab,kw.
63. "Logistic Model*".ti,ab,kw.
64. "clinical decision rule*".ti,ab,kw.
65. Stratification.ti,ab,kw.
66. exp receiver operating characteristic/
67. Discrimination.ti,ab,kw.
68. discriminate*.ti,ab,kw.
69. c-statistic.ti,ab,kw.

70. "c statistic".ti,ab,kw.
71. (Area adj under adj the adj curve).ti,ab,kw.
72. AUC.ti,ab,kw.
73. Calibration.ti,ab,kw.
74. Indices.ti,ab,kw.
75. Algorithm.ti,ab,kw.
76. Multivariable*.ti,ab,kw.
77. 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76
78. 53 and 77
79. limit 78 to (english language and yr="1984 -Current")

Notes

1. Predictive filter uses .ti,ab,kw rather than .mp as the database functionality is not as specific as MEDLINE.
2. There is no surgery subheading for the thesaurus term 'exp lumbosacral region/' therefore this is omitted.
3. There is no publication type available for case report or case study so unable to exclude.

A3. Cochrane CENTRAL

#1 MeSH descriptor: [Spine] explode all trees and with qualifier(s): [surgery - SU]
#2 (((spine or spinal) NEAR/5 (surger* or surgical*))):ti,ab,kw
#3 {OR #1-#2}
#4 MeSH descriptor: [Lumbar Vertebrae] explode all trees
#5 MeSH descriptor: [Low Back Pain] explode all trees
#6 MeSH descriptor: [Lumbosacral Region] explode all trees
#7 ("low* back" or lumbar or lumbosacral):ti,ab,kw
#8 {OR #4-#7}
#9 #3 AND #8
#10 MeSH descriptor: [Lumbosacral Region] explode all trees and with qualifier(s): [surgery - SU]
#11 MeSH descriptor: [Lumbar Vertebrae] explode all trees and with qualifier(s): [surgery - SU]
#12 MeSH descriptor: [Low Back Pain] explode all trees and with qualifier(s): [surgery - SU]
#13 MeSH descriptor: [Sciatica] explode all trees and with qualifier(s): [surgery - SU]
#14 (((lumb* or "lower back") NEAR/5 (surger* or surgical*))):ti,ab,kw
#15 {OR #10-#14}
#16 #9 OR #15
#17 MeSH descriptor: [Chronic Pain] explode all trees
#18 (((chronic or constant or persist* or longterm or long-term or "long standing" or longstanding or "long lasting" or longlasting) NEAR/5 (pain or lbp))):ti,ab,kw
#19 ((clbp or lumbar radiculopathy or lumbar radicular pain or sciatica or "postoperative pain")):ti,ab,kw

#20 MeSH descriptor: [Sciatica] explode all trees
#21 {OR #17-#20}
#22 #16 AND #21
#23 MeSH descriptor: [Pain] explode all trees
#24 MeSH descriptor: [Recovery of Function] explode all trees
#25 (pain*):ti,ab,kw
#26 ("Visual Analogue Scale"):ti,ab,kw
#27 (VAS):ti,ab,kw
#28 ("Numeric Rating Scale"):ti,ab,kw
#29 ("Numerical Rating Scale"):ti,ab,kw
#30 (NRS):ti,ab,kw
#31 ("Numeric Pain Rating Scale"):ti,ab,kw
#32 ("Numerical Pain Rating Scale"):ti,ab,kw
#33 (NPRS):ti,ab,kw

#34 (function*):ti,ab,kw
#35 (disabilit*):ti,ab,kw
#36 (ODI):ti,ab,kw
#37 ("Oswestry Disability Index"):ti,ab,kw
#38 ("Oswestry Disability Questionnaire"):ti,ab,kw
#39 ("Roland and Morris Disability Index"):ti,ab,kw
#40 ("Roland and Morris Disability Questionnaire"):ti,ab,kw
#41 (RMDQ):ti,ab,kw
#42 (RDQ):ti,ab,kw
#43 ("Core Outcome Measures Index"):ti,ab,kw
#44 (COMI):ti,ab,kw
#45 MeSH descriptor: [Return to Work] explode all trees
#46 ("return to work"):ti,ab,kw
#47 ("back to work"):ti,ab,kw
#48 ("work engagement"):ti,ab,kw
#49 (employment):ti,ab,kw
#50 {OR #23-#49}
#51 #22 AND #50
#52 MeSH descriptor: [Predictive Value of Tests] explode all trees
#53 MeSH descriptor: [Risk Factors] explode all trees
#54 Predict*:ti,kw
#55 prognos*:ti,kw
#56 Validat* or Rule*
#57 (Predict* and (Outcome* or Risk* or Model* or value))
#58 ((History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor*) and (Predict* or Model* or Decision* or Identif*OR Prognos*))
#59 (Decision* and (Model* or Clinical*))
#60 (Prognostic and (History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor* or Model* or value))
#61 MeSH descriptor: [Logistic Models] this term only
#62 MeSH descriptor: [Clinical Decision Rules] explode all trees
#63 Stratification
#64 MeSH descriptor: [ROC Curve] explode all trees
#65 Discrimination
#66 discriminate*
#67 c-statistic
#68 "c statistic"
#69 "Area under the curve"
#70 AUC
#71 Calibration
#72 Indices
#73 Algorithm
#74 Multivariable*
#75 {OR #52-#74}
#76 #51 AND #75
#77 MeSH descriptor: [Animals] this term only
#78 MeSH descriptor: [Humans] this term only
#79 #77 NOT #78
#80 #76 NOT #79 with Publication Year from 1984 to 2020, in Trials

Notes

1. Cannot limit to English language

A4. CINAHL

S1 (MH "Spine+/SU")

S2 TI ((spine or spinal) N5 (surger* or surgical*)) OR AB ((spine or spinal) N5 (surger* or surgical*))

S3 (MH "Lumbar Vertebrae/")
S4 (MH "Low Back Pain/")
S5 TI ("low* back" or lumbar or lumbosacral) OR AB ("low* back" or lumbar or lumbosacral)
S6 S1 OR S2
S7 S3 OR S4 OR S5
S8 S6 AND S7
S9 (MH "Lumbar Vertebrae/SU")
S10 (MH "Low Back Pain/SU")
S11 (MH "Sciatica/SU")
S12 TI ((lumb* or "lower back") N5 (surger* or surgical*)) OR AB ((lumb* or "lower back") N5 (surger* or surgical*))
S13 S9 OR S10 OR S11 OR S12
S14 S8 OR S13
S15 (MH "Chronic Pain")
S16 TI ((chronic or constant or persist* or longterm or long-term or "long standing" or longstanding or "long lasting" or longlasting) N5 (pain or lbp)) OR AB ((chronic or constant or persist* or longterm or long-term or "long standing" or longstanding or "long lasting" or longlasting) N5 (pain or lbp))
S17 TI (clbp or lumbar radiculopathy or lumbar radicular pain or sciatica or "postoperative pain") OR AB (clbp or lumbar radiculopathy or lumbar radicular pain or sciatica or "postoperative pain")
S18 (MH "Sciatica")
S19 S15 OR S16 OR S17 OR S18
S20 S14 AND S19
S21 (MH "Pain+")
S22 (MH "Functional Status")
S23 TI pain* OR AB pain*
S24 (MH "Visual Analog Scaling")
S25 TI "Visual Analogue Scale" OR AB "Visual Analogue Scale"
S26 TI VAS OR AB VAS
S27 TI "Numeric Rating Scale" OR AB "Numeric Rating Scale"
S28 TI "Numerical Rating Scale" OR AB "Numerical Rating Scale"
S29 TI NRS OR AB NRS
S30 TI "Numeric Pain Rating Scale" OR AB "Numeric Pain Rating Scale"
S31 TI "Numerical Pain Rating Scale" OR AB "Numerical Pain Rating Scale"
S32 TI NPRS OR AB NPRS
S33 TI function* OR AB function*
S34 TI disabilit* OR AB disabilit*
S35 TI ODI OR AB ODI
S36 TI "Oswestry Disability Index" OR AB "Oswestry Disability Index"
S37 TI "Oswestry Disability Questionnaire" OR AB "Oswestry Disability Questionnaire"
S38 TI ("Roland and Morris Disability Index") OR AB ("Roland and Morris Disability Index")
S39 TI ("Roland and Morris Disability Questionnaire") OR AB ("Roland and Morris Disability Questionnaire")
S40 TI RMDQ OR AB RMDQ
S41 TI RDQ OR AB RDQ
S42 TI "Core Outcome Measures Index" OR AB "Core Outcome Measures Index"
S43 TI COMI OR AB COMI
S44 (MH "Job Re-Entry")
S45 TI "return to work" OR AB "return to work"
S46 TI "back to work" OR AB "back to work"
S47 TI "work engagement" OR AB "work engagement"
S48 TI employment OR AB employment

S49 S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47 OR S48
S50 S20 AND S49
S51 (MH "Predictive Value of Tests")
S52 (MH "Risk Factors+")
S53 TI Predict*
S54 TI prognos*
S55 TI (Validat* or Rule*) OR AB (Validat* or Rule*)
S56 TI (Predict* and (Outcome* or Risk* or Model* or value)) OR AB (Predict* and (Outcome* or Risk* or Model* or value))
S57 TI ((History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor*) and (Predict* or Model* or Decision* or Identif* or Prognos*)) OR AB ((History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor*) and (Predict* or Model* or Decision* or Identif* or Prognos*))

S58 TI (Decision* and (Model* or Clinical*)) OR AB (Decision* and (Model* or Clinical*))
S59 TI (Prognostic and (History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor* or Model* or value)) OR AB (Prognostic and (History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor* or Model* or value))
S60 (MH "Multiple Logistic Regression")
S61 (MH "Decision Support Systems, Clinical")
S62 TI "clinical decision rule*" OR AB "clinical decision rule*"
S63 TI Stratification OR AB Stratification
S64 (MH "ROC Curve")
S65 TI Discrimination OR AB Discrimination
S66 TI discriminate* OR AB discriminate*
S67 TI c-statistic OR AB c-statistic
S68 TI "c statistic" OR AB "c statistic"
S69 TI "Area under the curve" OR AB "Area under the curve"
S70 TI AUC OR AB AUC
S71 TI Calibration OR AB Calibration
S72 TI Indices OR AB Indices
S73 TI Algorithm OR AB Algorithm
S74 TI Multivariable* OR AB Multivariable*
S75 S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62 OR S63 OR S64 OR S65 OR S66 OR S67 OR S68 OR S69 OR S70 OR S71 OR S72 OR S73 OR S74
S76 S50 AND S75
S77 PT Case Study
S78 S76 NOT S77
S79 (MH "Animals")
S80 (MH "Human")
S81 S79 NOT S80
S82 S78 NOT S81 Limiters - Publication Year: 1984-2020 Narrow by Language: - English

A5. PsycINFO

S1 DE "Spinal Column" OR DE "Spinal Cord"
S2 DE "Surgery"
S3 S1 AND S2
S4 TI ((spine or spinal) N5 (surger* or surgical*)) OR AB ((spine or spinal) N5 (surger* or surgical*))

S5 S3 OR S4

S6 DE "Lumbar Spinal Cord"
S7 TI ("low* back" or lumbar or lumbosacral) OR AB ("low* back" or lumbar or lumbosacral)
S8 S6 OR S7
S9 S5 AND S8
S10 TI ((lumb* or "lower back" or sciatica) N5 (surger* or surgical*)) OR AB ((lumb* or "lower back" or sciatica) N5 (surger* or surgical*))
S11 S9 OR S10
S12 DE "Chronic Pain"
S13 TI ((chronic or constant or persist* or longterm or long-term or "long standing" or longstanding or "long lasting" or longlasting) N5 (pain or lbp)) OR AB ((chronic or constant or persist* or longterm or long-term or "long standing" or longstanding or "long lasting" or longlasting) N5 (pain or lbp))
S14 TI (clbp or lumbar radiculopathy or lumbar radicular pain or sciatica or "postoperative pain") OR AB (clbp or lumbar radiculopathy or lumbar radicular pain or sciatica or "postoperative pain")
S15 S12 OR S13 OR S14
S16 S11 AND S15
S17 DE "Predictability (Measurement)"
S18 DE "Risk Factors"
S19 TI Predict*
S20 TI prognos*
S21 TI (Validat* or Rule*) OR AB (Validat* or Rule*)
S22 TI (Predict* and (Outcome* or Risk* or Model* or value)) OR AB (Predict* and (Outcome* or Risk* or Model* or value))
S23 TI ((History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor*) and (Predict* or Model* or Decision* or Identif* or Prognos*)) OR AB ((History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor*) and (Predict* or Model* or Decision* or Identif* or Prognos*))

S24 TI (Decision* and (Model* or Clinical*)) OR AB (Decision* and (Model* or Clinical*)) S25 TI (Prognostic and (History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor* or Model* or value)) OR AB (Prognostic and (History or Variable* or Criteria or Scor* or Characteristic* or Finding* or Factor* or Model* or value))
S26 TI "clinical decision rules*" OR AB "clinical decision rules*"
S27 TI ("ROC curve" OR stratification) OR AB ("ROC curve" OR stratification)
S28 TI Discrimination OR AB Discrimination
S29 TI discriminate* OR AB discriminate*
S30 TI c-statistic OR AB c-statistic
S31 TI "c statistic" OR AB "c statistic"
S32 TI ("Area under the curve" OR AUC) OR AB ("Area under the curve" OR AUC)
S33 TI Calibration OR AB Calibration
S34 TI Indices OR AB Indices
S35 TI Algorithm OR AB Algorithm
S36 TI Multivariable* OR AB Multivariable*
S37 S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36
S38 S16 AND S37
S39 S16 AND S37 Narrow by Language: - English

Notes

1. Earliest paper from 1986, so no date limits required
2. Small numbers retrieved without need to add in outcome terms

Supplemental Digital Content 2: Elaboration on eligibility criteria and decision rules

Population included adults with chronic low back pain, defined as pain that lasts or recurs for more than three months [80], with or without lumbar radicular pain, that is, pain radiating into the leg due to compression of a nerve root [81]. Where chronic pain population was not specified, or minimum pain duration was not reported, we allowed a minimum three months of failed conservative treatment as a proxy measure of symptom duration. We included patients who underwent primary lumbar or lumbosacral spine surgery, and excluded those with history of previous lumbar spine surgery, to distinguish chronic back and/or leg pain from failed back surgery syndrome and chronic postsurgical pain, which may differ in population characteristics and predictors of outcome [82-84]. For the same reasons, we excluded studies that reported on pathology of cancer/tumor, infection, trauma, or inflammatory disease. We further excluded spinal cord stimulator implantation, injections, chemical and radiofrequency interventions, as these interventions differ in indications and therapeutic mechanisms from spinal surgery. Studies specifically investigating whether additional pre- or postsurgical intervention (other than usual care) affects the outcomes were also excluded, as we aimed to investigate the outcomes of the surgical treatment itself.

Example *index prognostic factors* of interest included factors that are routinely used in patient assessment (e.g. 'flags' considered to be risk factors for the development of disability in musculoskeletal conditions [85]), have demonstrated ability to predict treatment outcomes in chronic pain in previous research (e.g. psychological factors, [86]), or have theoretical basis for such potential predictive ability (e.g. components of fear-avoidance model of chronic musculoskeletal pain [87]). We excluded radiographic predictors (as diagnostic tests or surgical indications), genetic predictors (due to their limited availability in clinical practice), and any intraoperative or postoperative predictors (beyond the scope of this review), unless they were included as *comparator prognostic factors*. There is no agreed minimum set of comparator prognostic factors for spinal surgery outcomes, thus both unadjusted and adjusted prognostic effects were eligible, where available.

Our decision to include both dichotomous (minimal clinically important difference in pain/disability) and continuous *outcomes* was motivated by their clinical utility and precision, respectively. We excluded studies that reported only postoperative assessment of pain or function without any baseline reference or used measures of pain or function that do not provide a continuous score as unsuitable for the assessment of change. We further excluded composite outcomes, as the predictive effects of the same factors might differ for each component outcome. The *timing* of outcome assessment must have been at least 3 months after surgery, without upper time limit. Where multiple follow-up intervals were analyzed, we included the last available follow-up, as commonly done in previous systematic reviews in the field [e.g., 84,92,93]. This solution also assured consistent treatment of all reviewed studies and maximized the use of available data as the last follow-up is often the main endpoint for which the primary analysis is conducted and therefore complete results are reported [e.g., 10,11,49,50,55,58]. Eligible *settings* included spinal surgery sites or registries / databases of patients who underwent spinal surgery.

We included a range of study designs (see **Table 1**), as long as they reported pre- and postsurgical assessment of outcomes and investigated relationships between presurgical factors and pain or function outcomes. Both prospective and retrospective studies were eligible – although retrospective investigations may be susceptible to poorer data quality and unmeasured predictors, they often allow analyzing longer follow-up and larger sample size than prospective studies, therefore providing valuable evidence in prognostic research. However, we excluded case studies and case series, which are considered to provide low level of evidence regarding prognosis [94,95].

We included peer-reviewed articles reporting original research published in English language from January 1984 to March 2021 (inclusive), whereas reviews, commentaries, editorial articles, conference abstracts, and study protocols were not eligible. Although we did not seek unpublished studies, we assessed publication bias on outcome level, and re-run the electronic database search prior to the final synthesis (March 2021) to include any recently published eligible articles. Non-English language publications were excluded as translation would not be feasible. The limits of the publication period reflect when Magnetic Resonance Imaging started being used for diagnosis of spinal pathologies [88] and thus could inform surgical treatments.

For 10 studies with insufficient details to determine eligibility, the necessary information was not available (n = 3) or could not be obtained from the authors (n = 7). However, five of these reports were included in the systematic review, because all other eligibility criteria were met and the information available or obtained from the authors suggested likely eligibility. This uncertainty was considered in the quality assessments. We excluded reports that lacked any indication of symptom duration in the studied cohorts. Reports that included subgroups with symptom duration <3 months were considered eligible if they presented analysis results allowing to extract data specific to eligible subgroups [54,60,10]. The same logic applied to reports that included subgroups undergoing surgical and conservative treatments if it was possible to extract data specific to eligible surgical cohorts [59-61,63].

The primary reason for exclusion was recorded as the first of the following categories for which eligibility was not met: ineligible population, predictors, outcome, timing, study design, publication type (see Figure 1). If eligibility in the former category could not be determined due to insufficient information reported, the subsequent exclusion category was recorded as the primary reason.

Supplemental Digital Content 3: Data extraction form template

Table S3. Data extraction form adapted from CHARMS-PF checklist [32].

		Source of data	**Participants**					
Reviewer	Study ID	Study design	Recruitment	Setting	Eligibility	Participant characteristics	Type of surgery	Participation
Initials, mark if verified	Author Year	Prospective / retrospective; RCT / non-randomized controlled trial / cohort study / case-control study / registry-based; phase of investigation	Method *(e.g. consecutive);* time period (years)	Number, type, and location of centers *(e.g. 5 Neurosurgery Departments in the UK)* / registry name	Inclusion and exclusion criteria (list)	Age (mean, SD); gender (% male); ethnicity (% each category); SES (% each category); pain location (% back, leg, both); pain intensity, disability (mean, SD); symptom duration (mean, SD; months); pathology (% each category)	% each category	% recruited out of those screened for eligibility

Sample size		**Prognostic factors (PF)**			
Calculation reported	N of participants	N of PFs	Definition, measurement, and handling of PFs	Time of PF measurement	Missing PF data
Yes / no; method *(e.g. events per variable)*	At baseline	Total N of index PFs (list); total N of comparator / adjusted for / confounding / controlled for PFs (list)	Construct measured, definition, method of measurement, *note if blinded, setting, handling of continuous factor (cut-off points or categories, if relevant), *note if not consistent across participants; for each index and comparator PF; author/date of non-standard measures	N of months / weeks / days / hours before surgery (pre-operative PFs)	% of sample with missing data on each PF, imputation method

Outcomes			**Attrition**		
Definition, measurement, and handling of outcomes	Time of outcome measurement	Missing outcome data	Response rate	Reasons for loss to follow-up	Lost to follow-up participant characteristics
Construct measured, definition, method of measurement, *note if blinded, setting, handling of continuous outcomes (cut-off points or categories, if relevant), *note if not consistent across participants; for each outcome (list non-eligible outcomes, but mark in red); author/year for non-standard measures	N of months / years after surgery; list all follow-ups	% of sample with missing data on each outcome, imputation method (or complete case analysis)	% of baseline sample with complete outcome data at the longest follow-up (unless eligible outcomes only available at earlier follow-up); method of handling missing data *(e.g. complete case analysis)*	% each category	Outcome / PF information collected (*note if not attempted); similar / different to completed participants

Analysis				
Modelling method	Modelling assumptions	Selection of PFs for multivariable modelling	Selection / exclusion of PFs during multivariable modelling	Handling continuous PFs
Univariate / multivariate; method *(e.g. linear, logistic, Cox, parametric survival, competing risks regression);* software	Which checked, what method, whether satisfied *(e.g. linearity, co-linearity for correlational/regression analyses; normality, homogeneity of variance, sphericity for ANOVA; parametric / non-parametric tests)*	Method *(e.g. all candidate PFs considered, preselection of established PFs, retaining those significant from univariable analysis);* selection criteria *(e.g. statistical significance)*	Method *(e.g. backward / forward selection, full model approach; order of entry if applicable);* criteria *(e.g. p, AIC)*	Method *(e.g. dichotomization, categorization, linear, non-linear);* any cut off with justification; method to identify non-linear associations *(e.g. splines, fractional polynomials)*

Results			
Effect estimates	Comparator PFs	Selective reporting	Interpretation and discussion
Adjusted and unadjusted effect estimates (RR / OR / HR / MD) with 95% CIs / variance / SE for each PF; for regression report F test with df and p, Rs, and final Bs and/or betas with SE, t-test and/or Ps for each PF, and mean with SD of outcome if available; N of participants included in each analysis if different from total sample size; format: by outcome	Set of adjusted for / confounding factors for each index PF (list)	Yes (list any primary outcomes and PFs included in methods but not results section) / no	Appropriate / inappropriate if missing results interpretation, comparison with other studies, discussion of generalizability and strengths and limitations

Supplemental Digital Content 4: Risk of bias assessment form template

Table S4. Adapted Quality of Prognosis Studies (QUIPS) [42–44] assessment form template.

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
Instructions to assess the risk of each potential bias:	These issues will guide your thinking and judgment about the overall risk of bias within each of the 6 domains. Some 'issues' may not be relevant to the specific study or the review research question. These issues are taken together to inform the overall judgment of potential bias for each of the 6 domains.	Provide comments or text excerpts in the boxes below, as necessary, to facilitate the consensus process that will follow.	Rating of reporting: Rate the adequacy of reporting as **yes**, **partial**, **no,** or **unsure**. Rating of RoB: Rate potential risk of bias for each of the 6 domains as **high**, **moderate**, or **low** considering all relevant issues.	Criteria to aid the rating decisions and interrater agreement (weights* in brackets).
1. Study Participation	*Goal: To judge the risk of selection bias (likelihood that relationship between PF and outcome is different for participants and eligible non-participants).*			
Source of target population	The source population or population of interest is adequately described for key characteristics (age, sex, socioeconomic status, duration of symptoms, location of pain, underlying pathology, type of surgery).		yes / partial / no / unsure	This item may not be taken into account since it is not common to report information of the source population in this field - this is rather covered by the eligibility criteria. (0)
Method used to identify population	The sampling frame and recruitment are adequately described, including methods to identify the sample sufficient to limit potential bias (number and type used,		yes / partial / no / unsure	**yes**: Information available on patients' recruitment, e.g. consecutive eligible patients, prospective or retrospective identification of eligible patients, etc. (1) *Note: Non-consecutive recruitment may lead to selective sampling bias.*

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
	e.g., referral patterns in health care).			
Recruitment period	Period of recruitment is adequately described.		yes / partial / no / unsure	**yes:** Information available in the beginning and end of recruitment or data collection, or the period covered in a database / registry search of eligible patients. (0.5)
Place of recruitment	Place of recruitment (setting and geographic location) are adequately described.		yes / partial / no / unsure	**yes:** Information available on the setting (or registry name), number and type of centers involved, and name of the hospital or geographical location (at least country). (1) *Note: Multi-center studies are likely to include more representative population than single-center ones.*
Inclusion and exclusion criteria	Inclusion and exclusion criteria are adequately described (e.g., including explicit diagnostic criteria or "zero time" description).		yes / partial / no / unsure	**yes:** At least 1 inclusion and 1 exclusion criterion should be given. (1)
Adequate study participation	There is adequate participation in the study by eligible individuals.		yes / partial / no / unsure	**yes:** Information available on number of individuals screened for eligibility / identified as eligible and approached (sample population) and those recruited into the study (study population), and reasons for exclusion; alternatively, participation should be at least 80%. (1) *Note: Data of non-included or lost participants may not be available, but quality can still be maintained if high participation / response rates are reported.*
Baseline characteristics	The baseline study sample (i.e., individuals entering the study) is adequately described for key characteristics (age, sex, socioeconomic status, duration of symptoms, location of pain, baseline pain intensity or disability,		yes / partial / no / unsure	**yes:** Basic information available regarding listed key characteristics (at least 6/8). (1)

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
	underlying pathology, type of surgery).			
Study Participation Summary	The study sample represents the population of interest on key characteristics, sufficient to limit potential bias of the observed relationship between PF and outcome.		**high / moderate / low**	**high:** The relationship between the PF and outcome is very likely to be different for participants and eligible non-participants. (1) ≥3 'no' / ≥2 'no' if the rest partial. *Example: low participation rate, study sample has different sex and age distribution than source population, recruitment of very selective rather than consecutive sample of eligible patients.* **moderate:** The relationship between the PF and outcome may be different for participants and eligible non-participants. (1) ≤1 'no' if the rest is 'yes' or 'partial' / ≤2 no if the rest is 'yes'. **low:** The relationship between the PF and outcome is unlikely to be different for participants and eligible non-participants. (1) All 'yes' with ≤2 'partial'. *Example: high participation rate, consecutive recruitment of eligible participants, sample characteristics similar to source population.*
2. Study Attrition	*Goal: To judge the risk of attrition bias (likelihood that the relationship between PF and outcome are different for completing and non-completing participants).*			
Proportion of baseline sample available for analysis	Response rate (i.e., proportion of study sample completing the study and providing outcome data) is adequate.		yes / partial / no / unsure	**yes:** The response rate should be at least 80% for the longest follow-up. (1) *Note: If response rate is 100% (could be assumed for registry data if only participants with complete follow-up were recruited), the remaining items not relevant (0). Responses to the last 3 items weighted according to response rate.*
Attempts to collect information on participants who dropped out	Attempts to collect information on participants who dropped out of the study are described.		yes / partial / no / unsure	**yes:** Information available on the methods and timing. (0.5) *Note: Unlikely to attempt collecting missing outcome data. If baseline data collected, it would be provided for the following items.*

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
Reasons and potential impact of subjects lost to follow-up	Reasons for loss to follow-up are provided.		yes / partial / no / unsure	**yes:** Any information available on the reasons for dropouts. (1) *Note: If reasons not consistent, it would suggest that participants were missing at random.*
Outcome and prognostic factor information on those lost to follow-up	Participants lost to follow-up are adequately described for key characteristics (age, sex, socioeconomic status, duration of symptoms, location of pain, baseline pain intensity or disability, underlying pathology, type of surgery).		yes / partial / no / unsure	**yes:** Basic information available regarding listed key characteristics (at least 4/8). (1) *Note: Unlikely that any outcomes would be collected from lost participants, focus on baseline characteristics (also below).*
	There are no important differences between key characteristics (see above) and outcomes in participants who completed the study and those who did not.		yes / partial / no / unsure	**yes:** There should be no clinically important or statistically significant differences between the completing participants and drop-outs regarding demographic and illness-related key characteristics, and outcomes if that information was collected. (1) *Note: 'no' if there are important differences; 'unsure' if not tested; 'partial' if differences on only some of the factors. If no differences, it would suggest that participants were missing at random.*
Study Attrition Summary	Loss to follow-up (from baseline sample to study population analyzed) is not associated with key characteristics (i.e., the study data adequately represent the sample) sufficient to limit potential bias to the observed relationship between PF and outcome.		**high / moderate / low**	**high:** The relationship between the PF and outcome is very likely to be different for completing and non-completing participants. (1) ≥2 'no' in the last 3 items if <80%. *Example: High probability that participants who completed the study and those who dropped out differ in a way that distorts the associations between predictors and outcomes.* **moderate:** The relationship between the PF and outcome may be different for completing and non-completing participants. (1) ≤1 'no' in the last 3 items if <80% / ≤3 no in the last 3 items if >80% / only complete cases eligible and no information on those lost to follow-up.

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
				low: The relationship between the PF and outcome is unlikely to be different for completing and non-completing participants. (1) 100% response rate / ≤3 'partial' in the last 3 items if >80%. *Example: Complete follow-up or evidence of participants missing at random.*
3. Prognostic Factor Measurement	*Goal: To judge the risk of measurement bias related to how PF was measured (differential measurement of PF related to the level of outcome).*			
Definition of the PF	A clear definition or description of 'PF' is provided (e.g., including dose, level, duration of exposure, and clear specification of the method of measurement).		yes / partial / no / unsure	**yes:** There should be a clear definition of PF, e.g. information on which questions were used, how the data was collected, or how the variable was constructed, etc. (1) *Note: What construct was measured with what instrument, total score or subscales used, self-reported or assessed by investigator.*
Valid and Reliable Measurement of PF	Method of PF measurement is adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall).		yes / partial / no / unsure	**yes:** There should be a reference of reliability / validity study or information on these features in the paper; when different prognostic factors are included with different RoB, this should be noted and solved in the data synthesis phase. (1) *Note: Note whether assessment was blinded / performed by independent investigator (not relevant for self-report); in prospective studies PF assessment is inherently blinded to outcome.*
	Continuous variables are reported or appropriate cut-points (i.e., not data-dependent) are used.		yes / partial / no / unsure	**yes:** Continuous used / any cut-offs used should NOT be based on the distribution of the data, but on established cut-offs in the field of chronic pain / spinal surgery. (1 / 0 if NA) *Note: Categorization of continuous factors contributes to loss of power, but main concern is how it was done; NA if factor initially categorical.*

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
Method and Setting of PF Measurement	The method and setting of measurement of PF is the same for all study participants.		yes / partial / no / unsure	**yes:** The PF should be the same, but the method or setting could be different provided that they are reliable (e.g. VAS or NRS); importantly, the timing of PF measurement relative to surgery should be reported and similar across participants. (1) *Note: ideally standardized assessment, performed in the same setting (home / hospital) by the same person.*
Proportion of data on PF available for analysis	Adequate proportion of the study sample has complete data for PF variable.		yes / partial / no / unsure	**yes:** There should be at least 80% available with complete data for any PF considered in the review. (1) *Note: 'unsure' if missingness not reported.*
Method used for missing data	Appropriate methods of imputation are used for missing 'PF' data.		yes / partial / no / unsure	**yes:** There should be some form of imputation used, but even if not, it could still be a 'yes' if 80% of the sample has complete data. (0.5) *Note: Imputation preferred, as complete case analysis might be invalid if data is not missing at random.*
PF Measurement Summary	*PF* is adequately measured in study participants to sufficiently limit potential bias.		high / moderate / low	**high:** The measurement of the PF is very likely to be different for different levels of the outcome of interest. (1) ≥2 'no'. *Example: Using unreliable methods of PF measurement, or different approaches for participants, which result in systematic misclassification.* **moderate:** The measurement of the PF may be different for different levels of the outcome of interest. (1) ≤1 'no'. **low:** The measurement of the PF is unlikely to be different for different levels of the outcome of interest. (1) ≤2 'partial' if the rest is 'yes'. *Example: PF measured similarly for all participants, using valid, reliable measures.*
4. Outcome Measurement	*Goal: To judge the risk of bias related to the measurement of outcome (differential measurement of outcome related to the baseline level of PF).*			

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
Definition of the Outcome	A clear definition of outcome is provided, including duration of follow-up and level and extent of the outcome construct.		yes / partial / no / unsure	**yes:** There should be a clear definition of outcome, e.g. information on which questions were used, how the data was collected, how the variable was constructed, whether any recommendations on outcome measures were used, etc.; timing of outcome measurement should be clearly stated. (1) *Note: What construct was measured with what instrument, total score or subscales used.*
Valid and Reliable Measurement of Outcome	The method of outcome measurement used is adequately valid and reliable to limit misclassification bias (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and confirmation of outcome with valid and reliable test).		yes / partial / no / unsure	**yes:** There should be a reference of reliability / validity study or information on these features in the paper; population on which reliability / validity was assessed should corresponds to the population of interest; blind measurement is not required as we're interested in patient-reported (pain, function) outcomes. (1)
Method and Setting of Outcome Measurement	The method and setting of outcome measurement is the same for all study participants.		yes / partial / no / unsure	**yes:** The outcome should be the same, but the method or setting could be different provided that they are reliable (e.g. VAS or NRS, in-person or phone / postal / online administration) and valid for the use in chronic pain / spinal surgery population; importantly, the timing of outcome measurement relative to surgery should be similar across participants. (1)
Outcome Measurement Summary	*Outcome of interest* is adequately measured in study participants to sufficiently limit potential bias.		**high / moderate / low**	**high:** The measurement of the outcome is very likely to be different related to the baseline level of the PF. (1) ≥2 'no'. *Example: Likely different measurement of outcome related to the extent of exposure to the prognostic factors.*

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
				moderate: The measurement of the outcome may be different related to the baseline level of the PF. (1) ≤1 'no'. **low:** The measurement of the outcome is unlikely to be different related to the baseline level of the PF. (1) ≤2 'partial' if the rest is 'yes'. *Example: outcome measured similarly for all participants, using valid, reliable measure.*
5. Study Confounding	*Goal: To judge the risk of bias due to confounding (i.e. the effect of PF is distorted by another factor that is related to PF and outcome).*			
Important Confounders Measured	All important confounders, including treatments (there is no specific set of required factors), are measured.		yes / partial / no / unsure	**yes:** There should be at least one confounder considered, as defined by the study authors; however, in a broad review in the field of CLBP with multifactor associations between prognostic factors and outcomes, it is not feasible to define a minimum set of potential confounders that should considered. (1) *Note: 'yes' if multivariate analysis.*
Definition of the confounding factor	Clear definitions of the important confounders measured are provided (e.g., including dose, level, and duration of exposures).		yes / partial / no / unsure	**yes:** There should be a clear definition of confounders, e.g. information on which questions were used, how the data was collected, or how the variable was constructed, etc. (1) *Note: What construct was measured with what instrument, total score or subscales used, self-reported or assessed by investigator; likely to be the same as for PFs if multivariate analysis.*
Valid and Reliable Measurement of Confounders	Measurement of all important confounders is adequately valid and reliable (e.g., may include relevant outside sources of information on measurement properties, also characteristics, such as blind measurement and limited reliance on recall).		yes / partial / no / unsure	**yes:** There should be a reference of reliability / validity study or information on these features in the paper (not relevant for basic demographic characteristics, e.g. sex or age); rationale for including a factor as a confounder should be provided. (1) *Note: Note whether assessment was blinded / performed by independent investigator (although not relevant for self-report); in prospective studies assessment is inherently blinded to outcome; likely to be the same as for PFs if multivariate analysis.*

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
Method and Setting of Confounding Measurement	The method and setting of confounding measurement are the same for all study participants.		yes / partial / no / unsure	**yes:** The confounder should be the same, but the method or setting could be different provided that they are reliable (e.g. VAS or NRS); importantly, the timing of confounder measurement relative to surgery should be reported and similar across participants. (1) *Note: Ideally standardized assessment, performed in the same setting (home / hospital) by the same person.*
Method used for missing data	Appropriate methods are used if imputation is used for missing confounder data.		yes / partial / no / unsure	**yes:** There should be some form of imputation used, but even if not, it could still be a 'yes' if 80% of the sample has complete data. (1) *Note: 'unsure' if missingness not reported.*
Appropriate Accounting for Confounding	Important potential confounders are accounted for in the study design (e.g., matching for key variables, stratification, or initial assembly of comparable groups).		yes / partial / no / unsure	**yes:** There should be some form of randomization, stratification, or matching for confounders in controlled studies. (1 / 0) *Note: This item and the one below treated interchangeably - either form of accounting for confounder(s) is sufficient for a 'yes'.*
	Important potential confounders are accounted for in the analysis (i.e., appropriate adjustment).			**yes:** Analysis should account for at least one confounder (i.e. multivariate analyses are assumed to be less biased, while univariate analyses are assumed to be associated with potential bias, although both are included in this review). (1 / 0) *Note: This item and the one above treated interchangeably - either form of accounting for confounder(s) is sufficient for a 'yes'; this item more likely if multivariate analysis.*
Study Confounding Summary	Important potential confounders are appropriately accounted for, limiting potential bias with respect to the relationship between *PF* and *outcome*.		**high / moderate / low**	**high:** The observed effect of the PF on the outcome is very likely to be distorted by another factor related to PF and outcome. (1) 'no' on the 1st item. *Example: Another factor related to both PF and the outcome is likely to explain the effect of the PF.* **moderate:** The observed effect of the PF on outcome may be distorted by another factor related to PF and outcome. (1) 'yes' on the 1st item and ≤2 'no' on the rest.

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
				low: The observed effect of the PF on outcome is unlikely to be distorted by another factor related to PF and outcome. (1) 'yes' on the 1st item and ≤3 partial on the rest. *Example: Adequate measurement of potential confounding variables and inclusion of these variables in a prespecified multivariate analysis.*
6. Statistical Analysis and Reporting	*Goal: To judge the risk of bias related to the statistical analysis and presentation of results.*			
Presentation of analytical strategy	There is sufficient presentation of data to assess the adequacy of the analysis.		yes / partial / no / unsure	**yes:** There should be enough information available to understand the statistical methods applied, so that it can be determined whether they are correct. (1) *Note: Tests should be specified, level of data (categorical / continuous); distinction between planned / follow-up analyses; sample size calculation or adequate N of events / participants (min. 10 events, e.g. achieving MCID, for dichotomous outcomes; min. 20 participants per continuous PF); significance level used.*
Model development strategy	The strategy for model building (i.e., inclusion of variables in the statistical model) is appropriate and is based on a conceptual framework or model.		yes / partial / no / unsure	**yes:** There should be information available on whether statistical assumptions were satisfied; if a model is described, methods of any pre-selection of variables and / or criteria for inclusion of variables in the model, with rationale, should be presented. (1) *Note: See data extraction table for possible approaches; large N of PFs in combination with small sample / events N increases risk of spurious correlations and overfitting; assumptions like normality or independence not relevant for correlational analyses; advised to check if non-linear transformations of continuous PFs are indicated; full model approach has lowest predictor selection bias but requires prior knowledge; stepwise / forward selection increases risk of overfitting; backward selection acceptable based on criteria like p / AIC / c-change.*
	The selected statistical model is adequate for the design of the study.		yes / partial / no / unsure	**yes:** There should be some form of statistical analysis description available, resulting in information on the effect of PF on the outcome. (1) *Note: CIs for effect estimates desired, exact p values.*

Biases	Issues to consider for judging overall rating of RoB	Study Methods & Comments	Rating	Criteria
Reporting of results	There is no selective reporting of results.		yes / partial / no / unsure	**yes:** All primary outcomes and PFs described in the method section should be included in the results section with words or in numbers (tables, figures). (1) *Note: Results for candidate predictors should be reported even if not significant; group / sample average descriptives for PFs and / or outcomes not directly relevant for correlational analysis but would allow to assess the prevalence of PF / outcome in the study sample.*
Statistical Analysis and Presentation Summary	The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid or spurious results.		**high / moderate / low**	**high:** The reported results are very likely to be spurious or biased related to analysis or reporting. (1) ≥2 'no'. *Example: Only significant results reported with omission of some primary outcomes and PFs.* **moderate:** The reported results may be spurious or biased related to analysis or reporting. (1) ≤1 'no' if the rest is 'yes' / 'partial'. **low:** The reported results are unlikely to be spurious or biased related to analysis or reporting. (1) ≤2 'partial' if the rest is 'yes'. *Example: Statistical analysis appropriate for the data, statistical assumptions satisfied, and all primary outcomes reported.*

Note. Current version of the form was adapted from QUIPS risk of bias assessment instrument for prognostic factor studies by Hayden et al. [44] based on original QUIPS version by Hayden et al. [43]. Criteria for ratings of reporting were adapted from Grooten et al. [42].

*Weights indicate the importance of each prompting item for determining the RoB rating in each bias domain (1 = important, 0 = not important)

Supplemental Digital Content 5: Excluded full text reports with reasons

1. Abbott AD, Tyni-Lenné R, Hedlund R. Leg pain and psychological variables predict outcome 2–3 years after lumbar fusion surgery. European Spine Journal 2011;20:1626–1634. **– Ineligible population (postoperative intervention) & Ineligible outcome (not change from baseline)**
2. Ablin JN, Berman M, Aloush V, Regev G, Salame K, Buskila D, Lidar Z. Effect of Fibromyalgia Symptoms on Outcome of Spinal Surgery. Pain medicine 2017;18:773–780. **– Ineligible population (cervical and lumbar spine surgery; symptom duration not reported) & Ineligible timing (short follow-up)**
3. Aghayev E, Roder C, Zweig T, Etter C, Schwarzenbach O. Benchmarking in the SWISSspine registry: results of 52 Dynardi lumbar total disc replacements compared with the data pool of 431 other lumbar disc prostheses. European spine journal 2010;19:2190–9. **– Ineligible population (symptom duration not recorded)**
4. Ahmadi SA, Burkert I-P, Steiger H-J, Eicker SO. Multidimensional long-term outcome analysis after single-level lumbar microdiscectomy: a retrospective single-centre study. European Journal of Orthopaedic Surgery & Traumatology 2018;28:189–196. **– Ineligible population (short symptom duration)**
5. Amundsen T, Weber H, Nordal HJ, Magnaes B, Abdelnoor M, Lilleas F. Lumbar spinal stenosis: Conservative or surgical management? A prospective 10-year study. Spine 2000;25:1424–1436. - **Ineligible outcome (composite/categorical measure) & Ineligible population (unclear symptom duration)**
6. Andersen MØ, Ernst C, Rasmussen J, Ankjær T, Carreon LY. Predictive Factors of Successful Return to Work Following Discectomy. Global Spine Journal 2020:2192568220960399. - **Ineligible population (short symptom duration)**
7. Andersen MO, Fritzell P, Eiskjaer SP, Lagerback T, Hagg O, Nordvall D, Lonne G, Solberg T, Jacobs W, van Hooff M, Gerdhem P, Gehrchen M. Surgical Treatment of Degenerative Disk Disease in Three Scandinavian Countries: An International Register Study Based on Three Merged National Spine Registers. Global Spine Journal 2019;9:850–858. **– Ineligible population (previous surgery)**
8. Andersen T, Christiansen FB, Laursen M, Høy K, Hansen ES, Bünger C. Smoking as a predictor of negative outcome in lumbar spinal fusion. Spine 2001;26:2623–2628. **– Ineligible population (previous surgery)**
9. Anderson JT, Haas AR, Percy R, Woods ST, Ahn UM, Ahn NU. Return to work after diskogenic fusion in workers' compensation subjects. Orthopedics 2015;38:e1065–e1072. **– Ineligible population (symptom duration not reported)**
10. Anderson JT, Haas AR, Percy R, Woods ST, Ahn UM, Ahn NU. Single-level lumbar fusion for degenerative disc disease is associated with worse outcomes compared with fusion for spondylolisthesis in a workers' compensation setting. Spine 2015;40:323–331. **– Ineligible population (previous surgery)**
11. Archer KR, Seebach CL, Mathis SL, Riley LH 3rd, Wegener ST. Early postoperative fear of movement predicts pain, disability, and physical health six months after spinal surgery for degenerative conditions. The spine journal 2014;14:759–67. **– Ineligible population (previous surgery, lumbar and cervical)**
12. Archer KR, Wegener ST, Seebach C, Song Y, Skolasky RL, Thornton C, Khanna AJ, Riley ILH. The effect of fear of movement beliefs on pain and disability after surgery for lumbar and cervical degenerative conditions. Spine 2011;36:1554–1562. **– Ineligible population (previous surgery, lumbar and cervical) & Ineligible study design (case series)**
13. Arpino L, Iavarone A, Parlato C, Moraci A. Prognostic role of depression after lumbar disc surgery. Neurological Sciences 2004;25:145–147. - **Ineligible study design (case series)**
14. Asher AL, Devin CJ, Archer KR, Chotai S, Parker SL, Bydon M, Nia, H, Harrell FE, Speroff T, Dittus RS, Philips SE, Shaffrey CI, Foley KT, McGirt MJ. An analysis from the Quality Outcomes Database, Part 2. Predictive model for return to work after elective surgery for lumbar degenerative disease. Journal of Neurosurgery: Spine 2017;27:370–381. **– Ineligible population (previous surgery, short symptom duration)**
15. Atlas SJ, Deyo RA, Keller RB, Chapin AM, Patrick DL, Long JM, Singer DE. The Maine Lumbar Spine Study, part II: 1-Year outcomes of surgical and nonsurgical management of sciatica. Spine 1996;21:1777–1786. **– Ineligible population (short symptom duration)**
16. Atlas SJ, Deyo RA, Keller RB, Chapin AM, Patrick DL, Long JM, Singer DE. The Maine Lumbar Spine Study, part III: 1-Year outcomes of surgical and nonsurgical management of lumbar spinal stenosis. Spine 1996;21:1787–1795. **– Ineligible population (short symptom duration)**

17. Atlas SJ, Keller RB, Chang Y, Deyo RA, Singer DE. Surgical and nonsurgical management of sciatica secondary to a lumbar disc herniation: Five-Year outcomes from the Maine lumbar spine study. Spine 2001;26:1179–1187. **– Ineligible population (short symptom duration)**
18. Behrend C, Prasarn M, Coyne E, Horodyski M, Wright J, Rechtine GR. Smoking Cessation Related to Improved Patient-Reported Pain Scores Following Spinal Care. The Journal of bone and joint surgery American volume 2012;94:2161–6. **– Ineligible population (no intervention)**
19. Bennett EE, Walsh KM, Thompson NR, Krishnaney AA. Central Sensitization Inventory as a Predictor of Worse Quality of Life Measures and Increased Length of Stay Following Spinal Fusion. World neurosurgery 2017;104:594–600. **– Ineligible population (thoracic and lumbar, symptom duration not reported)**
20. Berg S, Fritzell P, Tropp H. Sex life and sexual function in men and women before and after total disc replacement compared with posterior lumbar fusion. Spine journal 2009;9:987-994. **– Ineligible predictors (not baseline)**
21. Bernd L, Schiltenwolf M, Mau H, Schindele S. No indications for percutaneous lumbar discectomy? International Orthopaedics 1997;21:164–168. **– Ineligible population (age <18 years, short symptom duration)**
22. Bjarke Christensen F, Stender Hansen E, Laursen M, Thomsen K, Bünger CE. Long-term functional outcome of pedicle screw instrumentation as a support for posterolateral spinal fusion: randomized clinical study with a 5-year follow-up. Spine 2002;27:1269-1277. **Ineligible predictors (intraoperative) & Ineligible study design (no investigation of associations between preoperative factors and postoperative outcomes)**
23. Block AR, Ohnmeiss DD, Guyer RD, Rashbaum RF, Hochschuler SH. The use of presurgical psychological screening to predict the outcome of spine surgery. The spine journal: official journal of the North American Spine Society 2001;1:274–82. **– Ineligible population (previous surgery), Ineligible predictors (composite) & Ineligible outcome (composite)**
24. Blondel B, Tropiano P, Gaudart J, Huang RC, Marnay T, Blondel B, Tropiano P, Gaudart J, Huang RC, Marnay T. Clinical results of lumbar total disc arthroplasty in accordance with Modic signs, with a 2-year-minimum follow-up. Spine (03622436) 2011;36:2309–2315. **– Ineligible population (previous surgery)**
25. Bouras T, SStranjalis G, Loufardaki M, Sourtzis I, Stavrinou LC, Sakas DE. Predictors of long-term outcome in an elderly group after laminectomy for lumbar stenosis. Journal of Neurosurgery: Spine 2010;13:329–334. **– Ineligible population (previous surgery, unclear if lumbar, unclear symptom duration) & Ineligible study design (case series)**
26. Brox JI, Reikerås O, Nygaard Ø, Sørensen R, Indahl A, Holm I, Keller A, Ingebrigtsen T, Grundnes O, Lange JE, Friis A. Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: A prospective randomized controlled study. Pain 2006;122:145–155. **– Ineligible population (previous surgery)**
27. Burgstaller JM, Wertli MM, Steurer J, Kessels AGH, Held U, Gramke H-F, Group LS. The Influence of Pre- and Postoperative Fear Avoidance Beliefs on Postoperative Pain and Disability in Patients with Lumbar Spinal Stenosis: Analysis of the Lumbar Spinal Outcome Study (LSOS) Data. Spine 2017;42:E425–E432. **– Ineligible population (short symptom duration)**
28. Carreon LY, Jespersen AB, Støttrup CC, Hansen KH, Andersen MO. Is the Hospital Anxiety and Depression Scale Associated with Outcomes After Lumbar Spine Surgery? Global Spine Journal 2020;10:266–271. **– Ineligible population (symptom duration not reported) & Ineligible study design (case series)**
29. Carreon LY, Glassman SD, Kantamneni NR, Mugavin MO, Djurasovic M. Clinical outcomes after posterolateral lumbar fusion in workers compensation patients: A case-control study. Spine 2010;35:1812–1817. **– Ineligible population (previous surgery)**
30. Carreon LY, Glassman SD, Djurasovic M, Dimar JR, Johnson JR, Puno RM, Campbell MJ. Are Preoperative Health-Related Quality of Life Scores Predictive of Clinical Outcomes After Lumbar Fusion? Spine 2009;34. **– Ineligible population (previous surgery)**
31. Chaichana KL, Mukherjee D, Adogwa O, Cheng JS, McGirt MJ. Correlation of preoperative depression and somatic perception scales with postoperative disability and quality of life after lumbar discectomy. J Neurosurgery Spine 2011;14:261–7. **– Ineligible population (short symptom duration)**
32. Dance C, DeBerard MS, Cuneo JG. Pain acceptance potentially mediates the relationship between pain catastrophizing and post-surgery outcomes among compensated lumbar fusion patients. Journal of Pain

Research 2017;10:65–72. **– Ineligible predictors (not baseline), Ineligible outcome (not change from baseline) & Ineligible population (symptom duration not reported)**
33. D'Angelo C, Mirijello A, Ferrulli A, Leggio L, Berardi A, Icolaro N, Miceli A, D'Angelo V, Gasbarrini G, Addolorato G. Role of trait anxiety in persistent radicular pain after surgery for lumbar disc herniation: a 1-year longitudinal study. Neurosurgery 2010;67:265–71. **– Ineligible population (short symptom duration), Ineligible outcome (not change from baseline) & Ineligible study design (case series)**
34. De la Garza-Ramos R, Bydon M, Abt NB, Sciubba DM, Wolinsky J-P, Bydon A, Gokaslan ZL, Rabin B, Witham TF. The impact of obesity on short- and long-term outcomes after lumbar fusion. Spine 03622436 2015;40:56–61. **– Ineligible outcome (not change from baseline) & Ineligible population (symptom duration not reported)**
35. DeBerard MS, LaCaille RA, Spielmans G, Colledge A, Parlin MA. Outcomes and pre-surgery correlates of lumbar discectomy in Utah Workers' Compensation patients. Spine Journal 2009;9:193–203. **– Ineligible population (previous surgery, unclear symptom duration)**
36. den Boer JJ, Oostendorp RAB, Beems T, Munneke M, Evers AWM. Continued disability and pain after lumbar disc surgery: the role of cognitive-behavioral factors. Pain 2006;123:45–52. **– Ineligible population (short pain duration, age <18 years)**
37. Dipak S, Shrestha R, Dhoju D, Kayastha SR, Jha SC. Study of clinical variables affecting long term outcome after microdisectomy for lumbar disc herniation. Kathmandu University Medical Journal 2015;13:333–340. **– Ineligible population (short symptom duration) & Ineligible outcome (not change from baseline)**
38. Dzioba RB, Doxey NC. A prospective investigation into the orthopaedic and psychologic predictors of outcome of first lumbar surgery following industrial injury. Spine 1984;9:614–623. **– Ineligible population (short symptom duration, interventions included injections)**
39. Edwards RR, Klick B, Buenaver L, Max MB, Haythornthwaite JA, Keller RB, Atlas SJ. Symptoms of distress as prospective predictors of pain-related sciatica treatment outcomes. Pain 2007;130:47–55. **– Ineligible population (short symptom duration, surgical and conservative treatment)**
40. Ekselius L, Von Knorring L, Enskog J, Ordeberg G. Effect of personality disorders on treatment outcome of surgery for low back pain. Journal of Musculoskeletal Pain 1996;4:87–96. **– Ineligible predictors (not baseline)**
41. Ford JJ, Kaddour O, Page P, Richards MC, McMeeken JM, Hahne AJ. A multivariate prognostic model for pain and activity limitation in people undergoing lumbar discectomy. British Journal of Neurosurgery 2020.**– Ineligible population (short symptom duration)**
42. Franklin GM, Haug J, Heyer NJ, McKeefrey SP, Picciano JF. Outcome of lumbar fusion in Washington State workers' compensation. Spine (Phila Pa 1976) 1994;19:1897–903. **– Ineligible population (previous surgery, unclear symptom duration)**
43. Froholdt A, Reikeraas O, Holm I, Keller A, Brox JI. No difference in 9-year outcome in CLBP patients randomized to lumbar fusion versus cognitive intervention and exercises. European Spine Journal 2012;21:2531-2538. **– Ineligible population (previous surgery) & Ineligible study design (no investigation of associations between preoperative factors and postoperative outcomes)**
44. Furunes H, Hellum C, Brox JI, Rossvoll I, Espeland A, Berg L, Brogger HM, Smastuen MC, Storheim K. Lumbar total disc replacement: predictors for long-term outcome. European spine journal: official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society 2018;27:709–718. **– Ineligible population (previous surgery)**
45. Gehrchen MP, Dahl B, Katonis P, Blyme P, Tøndevold E, Kiær T. No difference in clinical outcome after posterolateral lumbar fusion between patients with isthmic spondylolisthesis and those with degenerative disc disease using pedicle screw instrumentation: a comparative study of 112 patients with 4 years of follow-up. European Spine Journal 2002;11:423–427. **Ineligible population (previous surgery, symptom duration not reported)**
46. Gepstein R, Shabat S, Arinzon ZH, Berner F, Catz A, Folman F. Does obesity affect the results of lumbar decompressive spinal surgery in the elderly? Clinical Orthopaedics and Related Research 2004:138–144. **– Ineligible study design (case series) & Ineligible population (symptom duration not reported)**
47. Giannadakis C, Nerland US, Solheim O, Jakola AS, Gulati M, Weber C, Nygaard EP, Solberg TK, Gulati S. Does Obesity Affect Outcomes after Decompressive Surgery for Lumbar Spinal Stenosis? A Multicenter,

Observational, Registry-Based Study. World Neurosurgery 2015;84:1227–1234. **– Ineligible population (symptom duration not recorded)**
48. Giannadakis C, Solheim O, Jakola AS, Nordseth T, Gulati AM, Nerland US, Nygaard ØP, Solberg TK, Gulati S. Surgery for Lumbar Spinal Stenosis in Individuals Aged 80 and Older: A Multicenter Observational Study. Journal of the American Geriatrics Society 2016;64:2011–2018. **– Ineligible population (previous surgery, unclear symptom duration)**
49. Graver V, Haaland AK, Magnaes B, Loeb M. Seven-year clinical follow-up after lumbar disc surgery: results and predictors of outcome. British Journal of Neurosurgery 1999;13:178–84. **– Ineligible population (short symptom duration)**
50. Graver V. New dissertation: long-term results and predictors of outcome in lumbar disc surgery with special attention to the impact of hypofibrinolysis, psychological distress and the extent of the surgical intervention. Advances in Physiotherapy 2000;2:93–95. **– Ineligible population (short symptom duration)**
51. Greenough CG, Taylor LJ, Fraser RD. Anterior lumbar fusion: Results, assessment techniques and prognostic factors. European Spine Journal 1994;3:225–230. **– Ineligible population (previous surgery, unclear symptom duration)**
52. Gulati S, Nordseth T, Nerland US, Gulati M, Weber C, Giannadakis C, Nygaard ØP, Solberg TK, Solheim O, Jakola AS. Does daily tobacco smoking affect outcomes after microdecompression for degenerative central lumbar spinal stenosis? – A multicenter observational registry-based study. Acta Neurochirurgica 2015;157:1157–1164. **– Ineligible population (previous surgery, symptom duration not recorded)**
53. Gum JL, Glassman SD, Carreon LY. Is Type of compensation a predictor of outcome after lumbar fusion? Spine 2013;38:443–448. **– Ineligible population (previous surgery, unclear symptom duration)**
54. Gunzburg R, Keller TS, Szpalski M, Vandeputte K, Spratt KF. Clinical and psychofunctional measures of conservative decompression surgery for lumbar spinal stenosis: A prospective cohort study. European Spine Journal 2003;12:197–204. **– Ineligible study design (case series) & Ineligible population (symptom duration not reported)**
55. Hakkinen A, Ylinen J, Kautiainen H, Airaksinen O, Herno A, Kiviranta I. Does the outcome 2 months after lumbar disc surgery predict the outcome 12 months later? Disability Rehabilitation 2003;25:968–72. **– Ineligible predictors (not baseline) & Ineligible outcomes (not change from baseline)**
56. Hanigan WC, Elwood PW, Henderson JP, Lister JR. Surgical results in obese patients with sciatica. Neurosurgery 1987;20:896–9. **– Ineligible population (short symptom duration)**
57. Hassan Hashisha RK, Ali SM, Awad TE. Impact of tramadol abuse on clinical outcome of lumbar discectomy patients'. NeuroToxicology 2019;75:9–13. **– Ineligible population (short symptom duration)**
58. Haugen AJ, Brox JI, Grovle L, Keller A, Natvig B, Soldal D, Grotle M. Prognostic factors for non-success in patients with sciatica and disc herniation. BMC Musculoskeletal Disorders 2012;13:183. **– Ineligible population (short symptom duration)**
59. Havakeshian S, Mannion AF. Negative beliefs and psychological disturbance in spine surgery patients: a cause or consequence of a poor treatment outcome? European Spine Journal 2013;22:2827–35. **– Ineligible population (additional pre/post-surgery intervention) & Ineligible outcomes (none of interest)**
60. Hellum C, Johnsen LG, Gjertsen Ø, Berg L, Neckelmann G, Grundnes O, Rossvoll I, Skouen JS, Brox JI, Storheim K. Predictors of outcome after surgery with disc prosthesis and rehabilitation in patients with chronic low back pain and degenerative disc: 2-year follow-up. European Spine Journal 2012;21:681-690. **– Ineligible population (previous surgery)**
61. Hennessy W, Wagner E, Dumas BP, Handley P. In Lumbar Fusion Patients, How Does Establishing a Comfort Function Goal Preoperatively Impact Postoperative Pain Scores? Pain Management Nursing 2015;16:841–845. **– Ineligible population (short symptom duration, additional pre/post-surgery intervention), Ineligible outcome (not change from baseline) & Ineligible timing (unclear follow-up duration)**
62. Herron LD, Turner J, Clancy S, Weiner P. The differential utility of the Minnesota multiphasic personality inventory: A predictor of outcome in lumbar laminectomy for disc herniation versus spinal stenosis. Spine 1986;11:847–850. **– Ineligible population (age <18 years, unclear pain duration) & Ineligible outcome (composite categorical)**
63. Hey HWD, Luo N, Chin SY, Lau ETC, Wang P, Kumar N, Lau LL, Ruiz JN, Thambiah JS, Liu KPG, Wong HK. The Predictive Value of Preoperative Health-Related Quality-of-Life Scores on Postoperative Patient-Reported

Outcome Scores in Lumbar Spine Surgery. Global Spine Journal 2018;8:156–163. **– Ineligible population (age <18 years, short symptom duration)**
64. Hodges SD, Humphreys SC, Eck JC, Covington LA, Harrom H. Predicting factors of successful recovery from lumbar spine surgery among workers' compensation patients. Journal of the American Osteopathic Association 2001;101:78–83. **– Ineligible population (short symptom duration)**
65. Hong X, Shi R, Wang Y-T, Liu L, Bao J-P, Wu X-T. Lumbar disc herniation treated by microendoscopic discectomy : Prognostic predictors of long-term postoperative outcome. Behandlung eines Bandscheibenvorfalls der Lendenwirbelsaule mittels endoskopischer Mikrodiskektomie : Prognostische Pradiktoren fur das postoperative Langzeitergebnis 2018;47:993–1002. **– Ineligible study design (case series)**
66. Hurri H, Slätis P, Soini J, Tallroth K, Alaranta H, Laine T, Heliövaara M. Lumbar spinal stenosis: Assessment of long-term outcome 12 years after operative and conservative treatment. Journal of Spinal Disorders 1998;11:110–115. **– Ineligible population (previous surgery, unclear symptom duration)**
67. Iversen MD, Daltroy LH, Fossel AH, Katz JN. The prognostic importance of patient pre-operative expectations of surgery for lumbar spinal stenosis. Patient Education and Counseling 1998;34:169–78. **– Ineligible outcomes (not change from baseline) & Ineligible population (symptom duration not reported)**
68. Jakobsson M, Brisby H, Gutke A, Hagg O, Lotzke H, Smeets R, Lundberg M. Prediction of Objectively Measured Physical Activity and Self-Reported Disability Following Lumbar Fusion Surgery. World Neurosurgery 2019;121:e77–e88. **– Ineligible population (previous surgery, additional pre-/post-surgery intervention)**
69. Jakola AS, Sørlie A, Gulati S, Nygaard ØP, Lydersen S, Solberg T. Clinical outcomes and safety assessment in elderly patients undergoing decompressive laminectomy for lumbar spinal stenosis: a prospective study. BMC Surgery 2010;10:34. **– Ineligible population (previous surgery, short symptom duration) & Ineligible study design (case series)**
70. Janssen ER, Osong B, Soest J van, Dekker A, Meeteren NL van, Willems PC, Punt IM. Exploring Associations of Preoperative Physical Performance With Postoperative Outcomes After Lumbar Spinal Fusion: A Machine Learning Approach. Archives of Physical Medicine and Rehabilitation 2021;102:1324-1330.e3. **– Ineligible population (previous surgery, unclear symptom duration)**
71. Jiménez-Almonte JH, Hautala GS, Abbenhaus EJ, Grabau JD, Nzegwu IN, Mehdi SK, Akhtar ZM, Liu B, Jacobs CA, Cassidy RC. Spine patients demystified: what are the predictive factors of poor surgical outcome in patients after elective cervical and lumbar spine surgery? Spine 2020;20:1529–1534. **– Ineligible study design (case series) & Ineligible population (symptom duration not recorded)**
72. Jensdottir M, G K , Hannesson, B , Gudmundsson, G. 20 Years follow-up after the first microsurgical lumbar discectomies in Iceland. Acta Neurochirurgica 2007;149:51–58. **– Ineligible population (age <18 years, symptom duration not reported)**
73. Johansson AC, Linton SJ, Rosenblad A, Bergkvist L, Nilsson O. A prospective study of cognitive behavioural factors as predictors of pain, disability and quality of life one year after lumbar disc surgery. Disability and Rehabilitation 2010;32:521–9. **– Ineligible population (short symptom duration) & Ineligible outcome (not change from baseline)**
74. Juratli SM, Franklin GM, Mirza SK, Wickizer TM, Fulton-Kehoe D. Lumbar fusion outcomes in Washington State workers' compensation. Spine 2006;31:2715–2723. **– Ineligible population (previous surgery, unclear symptom duration)**
75. Kerr D, Zhao W, Lurie JD. What Are Long-term Predictors of Outcomes for Lumbar Disc Herniation? A Randomized and Observational Study. Clinical Orthopaedics and Related Research 2015;473:1920–30. **– Ineligible population (short symptom duration)**
76. Khan JM, Basques BA, Harada GK, Louie PK, Chen I, Vetter C, Kadakia K, Elboghdady I, Colman M, An HS. Does increasing age impact clinical and radiographic outcomes following lumbar spinal fusion? Spine 2020;20:563–571. **– Ineligible population (previous surgery, unclear symptom duration) & Ineligible study design (case series)**
77. Khan JM, Harada GK, Basques BA, Nolte MT, Louie PK, Iloanya M, Tchalukov K, Berkowitz M, Derman P, Colman M, An HS. Patients with predominantly back pain at the time of lumbar fusion for low-grade spondylolisthesis experience similar clinical improvement to patients with predominantly leg pain: mid-

term results. Spine 2020;20:276–282. **Ineligible population (previous surgery, unclear symptom duration) & Ineligible study design (case series)**
78. Knutsson B, Michaëlsson K, Sandén B. Obesity is associated with inferior results after surgery for lumbar spinal stenosis: A study of 2633 patients from the Swedish spine register. Spine 2013;38:435–441. **– Ineligible population (previous surgery, symptom duration not reported)**
79. LaCaille RA, DeBerard MS, Masters KS, Colledge AL, Bacon W. Presurgical biopsychosocial factors predict multidimensional patient: outcomes of interbody cage lumbar fusion. Spine 2005;5:71–78. **– Ineligible population (previous surgery, unclear symptom duration) & Ineligible outcome (not change from baseline)**
80. Lagerbäck T, Fritzell P, Hägg O, Nordvall D, Lønne G, Solberg TK, Andersen MØ, Eiskjær S, Gehrchen M, Jacobs WC, van Hooff ML, Gerdhem P. Effectiveness of surgery for sciatica with disc herniation is not substantially affected by differences in surgical incidences among three countries: results from the Danish, Swedish and Norwegian spine registries. European Spine Journal 2019;28:2562–2571. **– Ineligible population (short symptom duration)**
81. Lequin MB, Verbaan D, Jacobs WCH, Brand R, Bouma GJ, Vandertop WP, Peul WC. Surgery versus prolonged conservative treatment for sciatica: 5-year results of a randomised controlled trial. BMJ Open 2013;3. **– Ineligible population (short symptom duration)**
82. Lewandrowski K-U, Ransom NA, Yeung A. Return to work and recovery time analysis after outpatient endoscopic lumbar transforaminal decompression surgery. Journal of Spine Surgery (Hong Kong) 2020;6:S100–S115. **– Ineligible study design (case series)**
83. Lewis PJ, Weir BK, Broad RW, Grace MG. Long-term prospective study of lumbosacral discectomy. Journal of Neurosurgery 1987;67:49–53. **- Ineligible study design (case series) & Ineligible population (symptom duration not reported)**
84. Lurie JD, Henderson ER, McDonough CM, Berven SH, Scherer EA, Tosteson TD, Tosteson ANA, Hu SS, Weinstein JN. Effect of Expectations on Treatment Outcome for Lumbar Intervertebral Disc Herniation. Spine (03622436) 2016;41:803–809. **– Ineligible population (short symptom duration)**
85. Madsbu MA, Oie LR, Salvesen O, Vangen-Lonne V, Nygaard OP, Solberg TK, Gulati S. Lumbar Microdiscectomy in Obese Patients: A Multicenter Observational Study. World Neurosurgery 2018;110:e1004–e1010. **– Ineligible population (symptom duration not reported)**
86. Madsbu MA, Salvesen O, Werner DAT, Franssen E, Weber C, Nygaard OP, Solberg TK, Gulati S. Surgery for Herniated Lumbar Disc in Daily Tobacco Smokers: A Multicenter Observational Study. World Neurosurgery 2018;109:e581–e587. **– Ineligible population (symptom duration not reported)**
87. Majeed SA, Vikraman CS, Mathew V, T. S A. Comparison of outcomes between conventional lumbar fenestration discectomy and minimally invasive lumbar discectomy: an observational study with a minimum 2-year follow-up. Journal of Orthopaedic Surgery & Research 2013;8:34–38. **– Ineligible population (short symptom duration)**
88. Mancuso CA, Reid MC, Duculan R, Girardi FP. Improvement in Pain After Lumbar Spine Surgery: The Role of Preoperative Expectations of Pain Relief. The Clinical Journal of Pain 2017;33:93–98. **– Ineligible population (previous surgery, unclear symptom duration)**
89. Mannion AF, Brox JI, Fairbank JC. Comparison of spinal fusion and nonoperative treatment in patients with chronic low back pain: long-term follow-up of three randomized controlled trials. Spine 2013;13:1438-1448. **– Ineligible population (previous surgery) & Ineligible study design (no investigation of associations between preoperative factors and postoperative outcomes)**
90. Marshman LAG, Kasis A, Krishna M, Bhatia CK. Does symptom duration correlate negatively with outcome after posterior lumbar interbody fusion for chronic low back pain? Spine 2010;35:657–65. **– Ineligible population (previous surgery), Ineligible outcome (not change from baseline) & Ineligible study design (case series)**
91. Mayo BC, Massel DH, Bohl DD, Long WW, Modi KD, Narain AS, Hijji FY, Lopez G, Singh K. Preoperative Mental Health is not Predictive of Patient-reported Outcomes Following a Minimally Invasive Lumbar Discectomy. Clinical Spine Surgery 2017;30:E1388–E1391. **– Ineligible population (symptom duration not recorded)**
92. McGirt MJ, Bydon M, Archer KR, Devin CJ, Chotai S, Parker SL, Nian H, Harrell FE Jr, Speroff T, Dittus RS, Philips SE, Shaffrey CI, Foley KT, Asher AL. An analysis from the Quality Outcomes Database, Part 1.

Disability, quality of life, and pain outcomes following lumbar spine surgery: predicting likely individual patient outcomes for shared decision-making. Journal of Neurosurgery: Spine 2017;27:357–369. **– Ineligible population (short symptom duration, previous surgery)**

93. McGuire KJ, Khaleel MA, Rihn JA, Lurie JD, Zhao W, Weinstein JN. The effect of high obesity on outcomes of treatment for lumbar spinal conditions: Subgroup analysis of the Spine Patient Outcomes Research Trial. Spine 2014;39:1975–1980. **– Other (duplicate of included report)**
94. Millisdotter M, Stromqvist B, Jonsson B. Proximal neuromuscular impairment in lumbar disc herniation: a prospective controlled study. Spine 2003;28:1281–9. **– Ineligible population (short symptom duration) & Ineligible outcome (not change from baseline)**
95. Moranjkic M, Zlatko E, Hodzic M, Brkic H. Outcome prediction in lumbar disc herniation surgery. Acta Medica Saliniana 2010;39:75–80. **– Ineligible population (short symptom duration)**
96. Müller M, Bütikofer L, Andersen OK, Heini P, Arendt-Nielsen L, Jüni P, Curatolo M. Cold pain hypersensitivity predicts trajectories of pain and disability after low back surgery: a prospective cohort study. PAIN 2021;162:184–194. **– Ineligible population (previous surgery) & Ineligible outcome (composite, trajectories)**
97. Mummaneni PV, Bisson EF, Kerezoudis P, Glassman S, Foley K, Slotkin JR, Potts E, Shaffrey M, Shaffrey CI, Coric D, Knightly J, Park P, Fu KM, Devin CJ, Chotai S, Chan AK, Virk M, Asher AL, Bydon M. Minimally invasive versus open fusion for Grade I degenerative lumbar spondylolisthesis: analysis of the Quality Outcomes Database. Neurosurgery Focus 2017;43:E11. - **Ineligible population (short symptom duration) & Ineligible predictors (intraoperative)**
98. Nickel R, Egle UT, Rompe J, Eysel P, Hoffmann SO. Somatisation predicts the outcome of treatment in patients with low back pain. Journal of Bone and Joint Surgery - Series B 2002;84:189–195. **– Ineligible population (previous surgery, surgical and conservative)**
99. Nygaard OP, Kloster R, Mellgren SI. Recovery of sensory nerve fibres after surgical decompression in lumbar radiculopathy: use of quantitative sensory testing in the exploration of different populations of nerve fibres. Journal of neurology, Neurosurgery, and Psychiatry 1998;64:120–3. **– Ineligible outcomes (none of interest), Ineligible study design (case series) & Ineligible population (unclear symptom duration)**
100. Olson PR, Lurie JD, Frymoyer J, Walsh T, Zhao W, Morgan TS, Abdu WA, Weinstein JN, Olson PR, Lurie JD, Frymoyer J, Walsh T, Zhao W, Morgan TS, Abdu WA, Weinstein JN. Lumbar disc herniation in the Spine Patient Outcomes Research Trial: does educational attainment impact outcome? Spine (03622436) 2011;36:2324–2332. **– Ineligible population (short symptom duration)**
101. Omidi-Kashani F, Hasankhani EG, Zare A. Prognostic value of impaired preoperative ankle reflex in surgical outcome of lumbar disc herniation. Archives of Bone and Joint Surgery 2016;4:52–55. **– Ineligible population (short symptom duration)**
102. Onyekwelu I, Glassman SD, Asher AL, Shaffrey CI, Mummaneni PV, Carreon LY. Impact of obesity on complications and outcomes: A comparison of fusion and nonfusion lumbar spine surgery. Journal of Neurosurgery: Spine 2017;26:158–162. **– Ineligible population (symptom duration not recorded)**
103. Parker SL, Godil SS, Zuckerman SL, Mendenhall SK, Devin CJ, McGirt MJ. Extent of preoperative depression is associated with return to work after lumbar fusion for spondylolisthesis. World Neurosurgery 2015;83:608–613. **– Ineligible population (short symptom duration)**
104. Petersen T, Juhl CB, Fournier GL. Patients With Persistent Low Back Pain and Nerve Root Involvement: To Operate, or Not to Operate, That Is the Question. Spine 2020;45:483–490. **– Ineligible population (previous surgery, short symptom duration) & Ineligible study design (case series)**
105. Plazier M, Raymaekers V, Bruyneel L, Coeckelberghs E, Sermeus W, Vanhaecht K, Duyvendak W. A 15-Year Follow-up Retrospective Study on 959 Spine Surgeries: What Can We Learn from Real-world Data? Clinical Spine Surgery 2021;34:E282. **– Ineligible outcome (none of interest) & Ineligible population (unclear symptom duration, age <18 years)**
106. Pollock R, Lakkol S, Budithi C, Bhatia C, Krishna M. Effect of psychological status on outcome of posterior lumbar interbody fusion surgery. Asian Spine Journal 2012;6:178–182. **– Ineligible population (previous surgery)**
107. Rihn JA, Kurd M, Hilibrand AS, Lurie J, Zhao W, Albert T, Weinstein J. The influence of obesity on the outcome of treatment of lumbar disc herniation: analysis of the Spine Patient Outcomes Research Trial

(SPORT). Journal of Bone and Joint Surgery American volume 2013;95:1-8. **– Ineligible population (short symptom duration)**
108. Riley IJL, Robinson ME, Geisser ME, Wittmer VT, Smith AG. Relationship between MMPI-2 cluster profiles and surgical outcome in low- back pain patients. Journal of Spinal Disorders 1995;8:213–219. **– Ineligible population (previous surgery)**
109. Sandén B, Försth P, Michaëlsson K. Smokers show less improvement than nonsmokers two years after surgery for lumbar spinal stenosis: A study of 4555 patients from the Swedish spine register. Spine 2011;36:1059–1064. **– Ineligible population (symptom duration not recorded)**
110. Seebach CL, Kirkhart M, Lating JM, Wegener ST, Song Y, Riley LH 3rd, Archer KR. Examining the role of positive and negative affect in recovery from spine surgery. PAIN 2012;153:518–25. **– Ineligible population (previous surgery, lumbar and cervical)**
111. Shamji MF, Shcharinsky A. Use of neuropathic pain questionnaires in predicting persistent postoperative neuropathic pain following lumbar discectomy for radiculopathy. Journal of Neurosurgery: Spine 2016;24:256–262. **– Ineligible study design (case series) & Ineligible population (symptom duration not reported)**
112. Siccoli A, Staartjes VE, de Wispelaere MP, Schröder ML. Association of time to surgery with leg pain after lumbar discectomy: is delayed surgery detrimental? Journal of Neurosurgery: Spine 2019:1–8. **– Ineligible population (short symptom duration)**
113. Siccoli A, Staartjes VE, de Wispelaere MP, Schröder ML. Association of time to surgery with leg pain after lumbar discectomy: is delayed surgery detrimental? Journal of Neurosurgery: Spine 2019:1–8. **– Ineligible population (short symptom duration)**
114. Siepe CJ, Mayer HM, Wiechert K, Korge A. Clinical results of total lumbar disc replacement with ProDisc II: Three-year results for different indications. Spine 2006;31:1923–1932. **– Ineligible study design (case series) & Ineligible population (symptom duration not reported)**
115. Silverplats K, Lind B, Zoëga B, Halldin K, Rutberg L, Gellerstedt M, Brisby H. Clinical factors of importance for outcome after lumbar disc herniation surgery: long-term follow-up. European Spine Journal 2010;19:1459–67. **– Ineligible study design (case series) & Ineligible population (unclear symptom duration)**
116. Sinigaglia R, Bundy A, Costantini S, Nena U, Finocchiaro F, Monterumici DAF. Comparison of single-level L4-L5 versus L5-S1 lumbar disc replacement: results and prognostic factors. European Spine Journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society 2009;18 Suppl 1:52–63. **– Ineligible population (previous surgery) & Ineligible predictors (intraoperative)**
117. Sinikallio S, Aalto T, Airaksinen O, Herno A, Kröger H, Savolainen S, Turunen V, Viinamäki H. Depression is associated with poorer outcome of lumbar spinal stenosis surgery. European Spine Journal 2007;16:905–912. **– Ineligible population (previous surgery)**
118. Sinikallio S, Aalto T, Koivumaa-Honkanen H, Airaksinen O, Herno A, Kröger H, Viinamäki H. Life dissatisfaction is associated with a poorer surgery outcome and depression among lumbar spinal stenosis patients: A 2-year prospective study. European Spine Journal 2009;18:1187–1193. **– Ineligible population (previous surgery)**
119. Sinikallio S, Aalto T, Airaksinen O, Herno A, Kröger H, Viinamäki H. Depressive burden in the preoperative and early recovery phase predicts poorer surgery outcome among lumbar spinal stenosis patients: a one-year prospective follow-up study. Spine (Phila Pa 1976) 2009;34:2573–8. **– Ineligible population (previous surgery, symptom duration not reported)**
120. Sinikallio S, Aalto T, Airaksinen O, Lehto SM, Kröger H, Viinamäki H. Depression is associated with a poorer outcome of lumbar spinal stenosis surgery: a two-year prospective follow-up study. Spine (Phila Pa 1976) 2011;36:677–682. **– Ineligible population (previous surgery)**
121. Sinikallio S, Lehto SM, Aalto T, Airaksinen O, Kröger H, Viinamäki H. Depressive symptoms during rehabilitation period predict poor outcome of lumbar spinal stenosis surgery: A two-year perspective. BMC Musculoskeletal Disorders 2010;11:152. **– Ineligible population (previous surgery)**
122. Skolasky RL, Scherer EA, Wegener ST, Tosteson TD. Does reduction in sciatica symptoms precede improvement in disability and physical health among those treated surgically for intervertebral disc herniation? Analysis of temporal patterns in data from the Spine Patient Outcomes Research Trial. The

Spine Journal : official journal of the North American Spine Society 2018;18:1318–1324. **– Ineligible population (short symptom duration) & Ineligible outcome (not change from baseline)**
123. Skolasky RL, Wegener ST, Maggard AM, Riley LH 3rd. The impact of reduction of pain after lumbar spine surgery: the relationship between changes in pain and physical function and disability. Spine 2014;39:1426–32. **– Ineligible predictors (not baseline) & Ineligible population (symptom duration not reported)**
124. Skouen JS, Larsen JL, Gjerde IO, Hegrestad SE, Vollset SE. Cerebrospinal fluid protein concentrations in patients with sciatica caused by lumbar disc herniation: an investigation of biochemical, neurologic, and radiologic predictors of long-term outcome. Journal of Spinal Disorders 1997;10:505–11. **– Ineligible outcomes (not change from baseline) & Ineligible population (symptom duration not reported)**
125. Slover J, Abdu WA, Hanscom B, Weinstein JN. The impact of comorbidities on the change in short-form 36 and Oswestry scores following lumbar spine surgery. Spine 2006;31:1974–1980. **– Ineligible population (symptom duration not reported)**
126. Solberg TK, Nygaard Øystein P, Sjaavik K, Hofoss D, Ingebrigtsen T. The risk of "getting worse" after lumbar microdiscectomy. European Spine Journal 2005;14:49–54. **– Ineligible study design (case series) & Ineligible population (unclear symptom duration)**
127. Sorensen LV, Mors O, Skovlund O. A prospective study of the importance of psychological and social factors for the outcome after surgery in patients with slipped lumbar disk operated upon for the first time. Acta neurochirurgica 1987;88:119–25. **– Ineligible outcomes (composite/categorical measure) & Ineligible population (unclear symptom duration and age)**
128. Staartjes VE, de Wispelaere MP, Vandertop WP, Schroder ML. Deep learning-based preoperative predictive analytics for patient-reported outcomes following lumbar discectomy: feasibility of center-specific modeling. The spine journal : official journal of the North American Spine Society 2019;19:853–861. **– Ineligible population (previous surgery, short symptom duration)**
129. Stienen MN, Smoll NR, Hildebrandt G, Schaller K, Gautschi OP. Influence of smoking status at time of surgery for herniated lumbar disk on postoperative pain and health-related quality of life. Clinical Neurology and Neurosurgery 2014;122:12–9. **– Ineligible population (previous surgery, symptom duration not reported) & Ineligible outcome (categorical measure)**
130. Tandon V, Campbell F, Ross ERS. Posterior lumbar interbody fusion: Association between disability and psychological disturbance in non-compensation patients. Spine 1999;24:1833–1838. **– Ineligible population (previous surgery)**
131. Triebel J, Snellman G, Sanden B, Stromqvist F, Robinson Y. Women do not fare worse than men after lumbar fusion surgery: Two-year follow-up results from 4,780 prospectively collected patients in the Swedish National Spine Register with lumbar degenerative disc disease and chronic low back pain. The Spine Journal: official journal of the North American Spine Society 2017;17:656–662. **– Ineligible population (previous surgery)**
132. Trief PM, G William, Fredrickson B. A Prospective Study of Psychological Predictors of Lumbar Surgery Outcome. Spine 2000;25. **– Ineligible population (previous surgery)**
133. Trief PM, Ploutz-Snyder R, Fredrickson BE. Emotional Health Predicts Pain and Function After Fusion: A Prospective Multicenter Study. Spine 2006;31:823–830. **– Ineligible population (previous surgery) & Ineligible outcome (not change from baseline)**
134. Tuna T, Boz S, Van Obbergh L, Lubansu A, Engelman E. Comparison of the Pain Sensitivity Questionnaire and the Pain Catastrophizing Scale in Predicting Postoperative Pain and Pain Chronicization After Spine Surgery. Clinical Spine Surgery 2018;31:E432–E440. - **Ineligible outcome (not change from baseline) & Ineligible population (symptom duration not reported)**
135. Tuomainen I, Pakarinen M, Aalto T, Sinikallio S, Kröger H, Viinamäki H, Airaksinen O. Depression is associated with the long-term outcome of lumbar spinal stenosis surgery: a 10-year follow-up study. The Spine Journal 2018;18:458–463. - **Ineligible outcome (not change from baseline)**
136. Tuomainen I, Pesonen J, Rade M, Pakarinen M, Leinonen V, Kröger H, Airaksinen O, Aalto T. Preoperative Predictors of Better Long-term Functional Ability and Decreased Pain Following LSS Surgery: A Prospective Observational Study with a 10-year Follow-up Period. Spine 2020;45:776–783. **– Ineligible population (previous surgery) & Ineligible outcome (not change from baseline)**

137. Turner JA, Herron L, Weiner P. Utility of the MMPI Pain Assessment Index in predicting outcome after lumbar surgery. Journal of Clinical Psychology 1986;42:764–9. **– Ineligible population (shorty symptom duration, age <18 years) & Ineligible outcome (not change from baseline, composite)**
138. Vaidya R, Carp J, Bartol S, Ouellette N, Lee S, Sethi A. Lumbar spine fusion in obese and morbidly obese patients. Spine 2009;34:495–500. **– Ineligible population (previous surgery)**
139. Van Susante J, Van De Schaaf D, Pavlov P. Psychological distress deteriorates the subjective outcome of lumbosacral fusion: A prospective study. Acta Orthopaedica Belgica 1998;64:371–377. **– Ineligible study design (case series)**
140. Vangen-Lønne V, Madsbu MA, Salvesen Ø, Nygaard ØP, Solberg TK, Gulati S. Microdiscectomy for Lumbar Disc Herniation: A Single-Center Observational Study. World Neurosurgery 2020;137:e577–e583. **– Ineligible population (age <18 years, unclear symptom duration)**
141. Vialle E, de Oliveira Pinto BM, Vialle LR, Gomez JDC. Evaluation of psychosomatic distress and its influence in the outcomes of lumbar fusion procedures for degenerative disorders of the spine. European Journal of Orthopopedic Surgery Traumatol 2015;25:25–28. **– Ineligible outcome (not change from baseline) &Ineligible population (symptom duration and age not reported)**
142. Vodičar M, Košak R, Gorenšek M, Korez R, Vrtovec T, Koder J, Antolič V, Vengust R. Vertebral End-Plate Perforation for Intervertebral Disc Height Preservation After Single-level Lumbar Discectomy: a Randomized-controlled Trial. Clinical Spine Surgery 2017;30:E707-E712 **– Ineligible population (short symptom duration) & Ineligible outcome (not change from baseline)**
143. Wahlman M, Häkkinen A, Dekker J, Marttinen I, Vihtonen K, Neva MH. The prevalence of depressive symptoms before and after surgery and its association with disability in patients undergoing lumbar spinal fusion. European Spine Journal 2014;23:129–134. **– Ineligible population (previous surgery)**
144. Wankhade UG, Umashankar MK, Jayakrishna Reddy BS. Functional outcome of lumbar discectomy by fenestration technique in lumbar disc prolapse-return to work and relief of pain. Journal of Clinical and Diagnostic Research 2016;10:9–13 **– Ineligible population (short symptom duration)**
145. Weinstein JN, T TD , Lurie, JD , Tosteson, ANA , Blood, E , Hanscom, B , Herkowitz, H , Cammisa, F , Albert, T , Boden, SD , Hilibrand, A , Goldberg, H , Berven, S , An, H. Surgical versus nonsurgical therapy for lumbar spinal stenosis. New England Journal of Medicine 2008;358:794–810. **– Ineligible predictors (intraoperative)**
146. Wilson-MacDonald J, Fairbank J, Frost H, Yu LM, Barker K, Collins R, Campbell H. The MRC spine stabilization trial: surgical methods, outcomes, costs, and complications of surgical stabilization. Spine 2008;33:2334-2340. **– Ineligible population (previous surgery), Ineligible predictors (intraoperative) & Ineligible study design (no investigation of associations between preoperative factors and postoperative outcomes)**
147. Yamashita K, Aono H, Yamasaki R. Clinical classification of patients with lumbar spinal stenosis based on their leg pain syndrome: Its correlation with 2-year surgical outcome. Spine 2007;32:980–985. **– Ineligible study design (Case series) & Ineligible population (symptom duration not reported)**
148. Ziegler DS, Jensen RK, Storm L, Carreon L, Andersen MO. The Association Between Early Postoperative Leg Pain Intensity and Disability at 1-Year and 2-Year Follow-Up After First-Time Lumbar Discectomy. Global Spine Journal 2021;11:81–88. **– Ineligible population (short symptom duration) & Ineligible outcome (not change from baseline)**

Supplemental Digital Content 6: Risk of bias judgements

Table S6. Risk of bias in the included studies.

	Bias Domains						Overall Assessment of Risk of Bias[b]
Study ID	**1. Study Participation**	**2. Study Attrition[a]**	**3. Prognostic Factor Measurement**	**4. Outcome Measurement**	**5. Study Confounding**	**6. Statistical Analysis and Reporting**	
Anderson 2006 [53]	high	moderate	low	moderate	low	moderate	HIGH
Cushnie 2019 [54]	moderate	high	moderate	low	low	moderate	HIGH
Ekman 2009 [55]	moderate	moderate	low	low	moderate	moderate	LOW
Gepstein 2007 [48]	moderate	moderate	moderate	low	moderate	high	HIGH
Hagg 2003 [59]	low	moderate	moderate	low	moderate	moderate	LOW
Hegarty 2012 [9]	low	low	low	low	low	moderate	LOW
Kim 2015a [49]	low	high	low	low	high	low	HIGH
Kim 2015b [50]	low	high	low	low	low	low	HIGH
McGuire 2014 [60]	moderate	high	low	low	moderate	low	HIGH
Muller 2019 [58]	moderate	moderate	moderate	low	low	moderate	LOW
Patel 2019 [51]	low	moderate	low	low	low	moderate	LOW
Radcliff 2011 [61]	moderate	high	moderate	low	moderate	moderate	HIGH
Rihn 2012 [62]	moderate	high	low	low	moderate	low	HIGH
Schade 1999 [63]	low	moderate	moderate	moderate	low	high	HIGH
Sigmundsson 2012 [52]	moderate	high	moderate	low	low	moderate	HIGH
Støttrup 2019 [56]	low	high	moderate	low	moderate	high	HIGH
Stromqvist 2008 [57]	moderate	moderate	low	low	high	high	HIGH
Tampin 2020 [64]	moderate	moderate	low	low	low	moderate	LOW
Watkins 1986 [65]	high	moderate	low	moderate	moderate	moderate	HIGH
Zweig 2011 [11]	low	moderate	moderate	moderate	low	moderate	LOW
Zweig 2017 [10]	low	high	moderate	moderate	low	moderate	HIGH

[a] Study attrition ratings were downgraded from low to moderate for studies that retrospectively included only those patients who had complete follow-up data and did not provide information about the number and/or characteristics of patients who were not included due to incomplete follow-up data [11,51,57,65].

[b] Low = all domains low or moderate; high = one or more domains high [45].

Supplemental Digital Content 7: Results of syntheses and discussion of non-predictors of pain and disability outcomes

1. Results

1.1. Primary outcome: change in pain intensity

1.1.1. Sociodemographic factors

Four studies (1 low, 3 high RoB) examined the prognostic effect of *gender*. Female gender was associated with better pain outcomes in both ethnic cohorts in 1 high RoB study (small effect sizes in unadjusted analyses) [48], whereas the remaining studies reported no significant associations in adjusted [53] and unadjusted analyses [9,57].

Only a single high RoB study [48] investigated the potential effect of *ethnicity* (Israeli Arabs vs. Israeli Jews) in unadjusted analysis, indicating no significant association with pain outcomes.

A single phase-2 high RoB study [53] evaluated the independent effect of pre-operative *work status* on pain outcomes, which was not significant. Another phase-1 low RoB study [55] suggested that patients who were working before surgery had greater improvement in pain than those not working, however, data supporting this conclusion could not be extracted and thus it was not included in the current review.

One high RoB study [53] assessed the prognostic value of *worker's compensation status*, which had no significant effect in adjusted analysis.

1.1.2. Health-related factors

One low RoB study [9] tested the effect of *pain quality* on change in pain after and found no significant association in unadjusted analysis.

One low RoB study [9] assessed the prognostic value of *sensory detection threshold* for pain outcomes, demonstrating no significant association in unadjusted analysis. Two low RoB studies [9,58] examined the effects of *pain detection thresholds*, which consistently did not predict pain outcomes in adjusted [58] and unadjusted [9] analyses. The same 2 studies found no effect of *pain tolerance thresholds* on pain outcomes in adjusted [58] and unadjusted analyses [9].

A single low RoB study [58] assessed the effect of *conditioned pain modulation* on pain outcomes, reporting no significant association in adjusted analysis.

Four studies (2 low RoB, 2 high RoB) examined the effect of baseline *disability* on pain outcomes. One low RoB study [9] and 1 high RoB study [52] reported that lower disability defined as RMDQ score or self-reported walking distance of more than 1000 meters was associated with better pain outcomes (small and unclear effects, respectively) in adjusted analyses, but not in unadjusted analyses [9]. The 2 remaining studies, including 1 phase-2 study [51], found no association between these factors in adjusted analyses [51,53]. The discrepancy in the results may be due to different definition of outcome and shorter follow-up time (70% pain reduction 3 months after surgery) in Hegarty et al.'s [9] study compared to other studies, and different definition of the prognostic factor in Sigmundsson et al.'s study (walking distance) [52].

Only a single, high RoB study assessed the effect of *smoking* status on pain outcomes, which was not significant in an adjusted analysis [53].

1.1.3. Psychological factors

A single low RoB study assessed the effect of *mental functioning* on pain outcomes, reporting no significant association in unadjusted analysis [9].

A single high RoB study investigated the effect of *pain sensitivity* score on reduction of back and leg pain in unadjusted analyses, reporting no significant associations with either outcome [49].

One low RoB study assessed the prognostic value of *pain drawing*, reporting no significant association of organic vs. non-organic signs with pain outcomes in unadjusted analysis [55].

1.2. Secondary outcome: change in disability

1.2.1. Sociodemographic factors

A single high RoB study examined the effect of *age* categorized as more or less than 48 years on disability outcome. Adjusted analysis showed no significant association [53].

Three high RoB studies assessed the prognostic value of *gender*. One adjusted [53] and 2 unadjusted [57,65] analyses consistently indicated that gender was not related to disability outcomes.

One high RoB study compared the change in disability between Israeli Arabs and Israeli Jews, showing no significant effect of *ethnicity* in unadjusted analysis [48].

A single high RoB phase-2 study assessed the independent contribution of pre-operative *work status* on disability outcomes in adjusted analysis, which showed no significant effect of this factor. Another phase-1 low RoB study [3] indicated that patients who were working before surgery had greater improvement in disability than those not working, however, this finding was not included in the current review because it was not possible to extract any supporting data.

One high RoB study indicated that *worker's compensation* claim also did not significantly affect the disability outcome in adjusted analysis.

1.2.2. Health-related factors

One low RoB study assessed whether having *sciatica* (measured using a Pain Drawing) affected disability outcomes, showing no significant effect in unadjusted analysis [55].

Two studies (1 low, 1 high RoB) assessed the effect of preoperative *pain intensity* on disability outcome, indicating no significant associations in adjusted [53] and unadjusted [64] analyses. There was one exception, where out of a range of pain-related candidate predictors including average leg and back pain intensity and bothersomeness in the past 24 hours or 1-2 weeks and neuropathic pain component score, only average leg pain intensity over last week demonstrated a large marginally significant difference between patients who achieved MCID in disability (higher preoperative pain intensity) and those who did not [64].

Two phase-2 high RoB reports based on the same study investigated whether *body mass index* is an independent predictor of disability outcomes in spinal stenosis and degenerative spondylolisthesis cohorts. Both reports demonstrated no significant association between obesity and disability outcomes in adjusted analyses in either patient cohort, regardless of whether the body mass index was categorized into 'no obesity' and 'obesity' [62] or included an additional 'extreme obesity' category [60]. One exception was that non-obese relative to obese patients reported greater improvement on SF-36 PF (small effect), but not on ODI [62].

A single high RoB study assessed the effect of *smoking* status on disability outcome, indicating no significant association in adjusted analysis [53].

One low RoB study [64] examined the effect of *sleep quality* on disability outcomes, reporting no significant difference in unadjusted analysis.

1.2.3. Psychological factors

Two studies (1 low, 1 high RoB) [50,64] examined the effect of *pain catastrophizing* on disability outcomes in unadjusted analyses. Classification of patients into low and high pain catastrophizing

did significantly interact with changes in ODI across three follow-up time points, indicating greater reduction in disability in high pain catastrophizing group, however, the difference in how much each group improved from baseline to 12 months follow-up was not statistically significant [50]. Similarly, average pain catastrophizing scores did not differ between patients who did and those who did not achieve MCID in disability [64].

One low RoB study [64] assessed the effect of *kinesiophobia* (fear of pain due to movement) on disability outcomes, showing no significant differences in unadjusted analysis.

An effect of *pain sensitivity* on disability outcome was assessed in one high RoB study [49]. There were no significant differences between low and high pain sensitivity groups in the degree of ODI improvement in unadjusted analyses.

A single low RoB study [64] assessed the effect of *mental functioning* on disability outcomes, reporting no significant association in unadjusted analysis.

The effect of *anxiety* on achieving MCID in disability was examined in one low RoB study [64], reporting no significant association.

2. Discussion

While younger age and higher education level are potential predictors of greater reductions in pain after surgery, we found no evidence for prognostic value of other *sociodemographic* characteristics, that is, gender and work-related factors, for either pain or disability outcomes.

For a range of *health-related* factors, we found no associations with reduction in pain or disability outcomes. Presence of sciatica in spondylolisthesis was unrelated to pain or disability outcomes (very low-quality evidence), although there was some evidence that it may predict pain outcomes in spinal stenosis, which was generally associated with less reduction in pain after surgery. Apart from the likely associations between greater sensory loss and more improvement in disability, the current review provides very low-quality evidence that quantitative sensory testing, including pressure pain sensitivity and tolerance [58], may be unrelated to change in pain intensity after surgery (although e.g. enhanced temporal summation, consistently predicting persistent postsurgical pain [89], has not yet been investigated in spinal surgery context). While there may be an effect of baseline disability on disability outcomes, we found very low-quality evidence for no association with change in pain intensity after surgery. Contrary to the assumption that obesity or smoking would be related to worse health outcomes more generally, there is no sufficient evidence to suggest that these factors should inform selection of patients with chronic LBP for surgery. While sleep disturbance has been linked to the development and severity of chronic pain symptoms [90,91], we found very low-quality evidence that sleep quality is unrelated to change in disability after surgery [64].

Not all *psychological* factors demonstrated predictive ability for reduction in pain or disability after surgery. Although pain-related factors, driven by pain catastrophizing, were associated with pain outcomes, pain drawing and pain sensitivity had limited prognostic value. The same factors as well as kinesiophobia were unrelated to disability outcomes, possibly because they represent pain-specific psychological constructs. In contrast to its association with pain outcomes, anxiety was not related to improvement in disability. General mental functioning showed no relationships with either surgical outcome. Notably, majority of these psychological factors were only tested in single studies.

Supplemental Digital Content 8: GRADE quality of evidence assessment

Table S8a. Quality of evidence for associations between baseline prognostic factors and change in pain intensity outcome.

Potential PFs	N of reports; cohorts; participants	Unadjusted			Adjusted			Phase	GRADE factors							Overall quality
									Downgrade if ✕					Upgrade if ✓		
		+	0	-	+	0	-		Study limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Mod./large effect size	Dose effect	
Sociodemographic																
DEMOGRAPHIC (age, gender, ethnicity)	4 reports [9,48,53,57]; 5 cohorts; 680 participants		2	2		2	1	1	✕	✓	✓	✕	✓	✕	✓	++
SOCIOECONOMIC (education, work status, worker's compensation)	2 reports [48,53]; 3 cohorts; 326 participants	1	1		1	1		2	✕	✓	✓	✓	✕	✕	✕	++
Health-related																
DIAGNOSIS (spinal pathology, sciatica)	4 reports [11,48,52,55]; 5 cohorts; 893 participants	1	3		2			1	✓	✓	✓	✕	✓	?	✕	++
SYMPTOM DURATION	4 reports [9,10,54,56]; 4 cohorts; 4066 participants		3	1		1		2	✕	✓	✓	✓	✓	✕	✕	+++
PAIN (pain intensity, pain quality, night-time pain)	3 reports [9,48,53]; 4 cohorts; 379 participants		2	2		1	1	1	✕	✓	✕	✕	✕	✕	✕	+
QST (sensory detection, pain detection, and pain tolerance threshold, CPM)	2 reports [9,58]; 2 cohorts; 116 participants		2			2		1	✓	✓	✓	✕	✕	✕	✕	+
DISABILITY	4 reports [9,51–53]; 4		1			2	2	1	✓	✓	✕	✕	✓	✕	✕	+

Potential PFs	N of reports; cohorts; participants	Unadjusted			Adjusted			Phase	GRADE factors							Overall quality
									Downgrade if ✕					Upgrade if ✓		
		+	0	-	+	0	-		Study limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Mod./large effect size	Dose effect	
	cohorts; 318 participants															
COMOBRBIDITIES (comorbidity, BMI, smoking)	2 reports [48,53]; 3 cohorts; 326 participants			2		1	2	1	✕	✓	✓	✕	✕	✓	✕	+
Psychological																
PAIN-RELATED (pain catastrophizing, pain sensitivity, pain drawing)	4 reports [9,49,50,55]; 3 cohorts; 388 participants	1	3	1				1	✕	✓	✕	✕	✓	?	?	+
AFFECTIVE (, mental functioning, anxiety, depression)	2 reports [9,48]; 3 cohorts; 273 participants		1	2				1	✕	✓	✕	✕	✕	✓	?	+

Phase, phase of investigation determining the starting quality of evidence before downgrading/upgrading based on GRADE factors (phase-1, moderate; phase-2, high). For unadjusted and adjusted analyses: '+', number of significant effects with a positive value (presence of or higher score on the prognostic factor is associated with better outcome, or absence of or lower score with worse outcome); '0', number of not significant effects; '-', number of significant effects with a negative value (absence of or lower score on the prognostic factor is associated with better outcome, or presence of or higher score with worse outcome); where multiple analyses per cohort per predictor were reported, only the most robust/representative one was considered in GRADE; unadjusted and adjusted effects were counted separately. For GRADE factors: ✓, no serious limitations (or present for moderate/large effect size, dose effect); ✕, serious limitations (or not present for moderate/large effect size, dose effect); ?, unable to rate item based on available information; NA, not applicable. For overall quality of evidence: +, very low; ++, low; +++, moderate; ++++, high.

Table S8b. Quality of evidence for associations between baseline prognostic factors and change in disability outcome.

Potential PFs	N of reports; cohorts; participants	Unadjusted			Adjusted			Phase	GRADE factors							Overall quality
									Downgrade if ×					Upgrade if ✓		
		+	0	-	+	0	-		Study limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Mod./large effect size	Dose effect	
Sociodemographic																
DEMOGRAPHIC (age, gender, ethnicity)	4 reports [48,53,57,65]; 4 cohorts; 656 participants		4			1		1	×	✓	✓	×	✓	×	NA	+
SOCIOECONOMIC (work status, worker's compensation)	1 report [53]; 1 cohort; 93 participants					1		2	×	NA	✓	✓	×	×	NA	++
Health-related																
DIAGNOSIS (sciatica)	1 report [55]; 1 cohort; 164 participants		1					1	×	NA	✓	×	×	×	NA	+
SYMPTOM DURATION	5 reports [10,54,56,61,64]; 6 cohorts; 4474 participants		3	1		2	1	2	×	×	✓	✓	✓	×	×	++
PAIN (pain intensity, bothersomeness, neuropathic component)	2 reports [53,64]; 2 cohorts; 141 participants		1			1		1	×	✓	✓	✓	×	×	×	+
QST (sensory detection threshold)	1 report [64]; 1 cohort; 48 participants	1			1			2	✓	NA	✓	×	×	✓	NA	+++
DISABILITY	4 reports [51,53,63,64]; 4 cohorts; 313 participants		1			1	2	2	✓	×	✓	×	✓	×	×	++
COMORBIDITIES (BMI, smoking)	3 reports [53,60,62]; 3					3		2	×	✓	✓	✓	×	×	×	++

Potential PFs	N of reports; cohorts; participants	Unadjusted			Adjusted			Phase	GRADE factors							Overall quality
									Downgrade if ✕					Upgrade if ✓		
		+	0	-	+	0	-		Study limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Mod./large effect size	Dose effect	
	cohorts; 897 participants															
SLEEP	1 report [64]; 1 cohort; 48 participants		1					1	✓	NA	✓	✕	✕	✕	NA	+
Psychological																
PAIN-RELATED (pain catastrophizing, pain sensitivity, pain drawing, kinesiophobia)	4 reports [49,50,55,64]; 3 cohorts; 383 participants		2	1				1	✕	✓	✕	✕	✓	✕	✕	+
AFFECTIVE-MOTIVATIONAL (mental functioning, anxiety, depression, vitality, job-related resignation)	2 reports [63,64]; 2 cohorts; 90 participants		1	1			1	1	✕	✓	✓	✕	✕	✓	NA	+
PERSONALITY (neuroticism)	1 report [59]; 1 cohort; 183 participants			1				1	✕	NA	✓	✕	✕	✕	NA	+

Phase, phase of investigation determining the starting quality of evidence before downgrading/upgrading based on GRADE factors (phase-1, moderate; phase-2, high). For unadjusted and adjusted analyses: '+', number of significant effects with a positive value (presence of or higher score on the prognostic factor is associated with better outcome, or absence of or lower score with worse outcome); '0', number of not significant effects; '-', number of significant effects with a negative value (absence of or lower score on the prognostic factor is associated with better outcome, or presence of or higher score with worse outcome); where multiple analyses per cohort per predictor were reported, only the most robust/representative one was considered in GRADE; unadjusted and adjusted effects were counted separately. For GRADE factors: ✓, no serious limitations (or present for moderate/large effect size, dose effect); ✕, serious limitations (or not present for moderate/large effect size, dose effect); ?, unable to rate item based on available information; NA, not applicable. For overall quality of evidence: +, very low; ++, low; +++, moderate; ++++, high.

References

1. Cited in the main article text (matched reference numbers)

9. Hegarty D, Shorten G. Multivariate prognostic modeling of persistent pain following lumbar discectomy. Pain Physician 2012;15(5):421–34.
10. Zweig T, Enke J, Mannion AF, Sobottke R, Melloh M, Freeman BJ, et al. Is the duration of pre-operative conservative treatment associated with the clinical outcome following surgical decompression for lumbar spinal stenosis? A study based on the Spine Tango Registry. Eur Spine J 2016/12/17 edn 2017;26(2):488–500.
11. Zweig T, Hemmeler C, Aghayev E, Melloh M, Etter C, Röder C. Influence of preoperative nucleus pulposus status and radiculopathy on outcomes in mono-segmental lumbar total disc replacement: results from a nationwide registry. BMC Musculoskelet Disord 2011;12(1):275–275.
32. Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ 2019/02/01 edn 2019;364:k4597.
42. Grooten WJA, Tseli E, Ang BO, Boersma K, Stalnacke BM, Gerdle B, et al. Elaborating on the assessment of the risk of bias in prognostic studies in pain rehabilitation using QUIPS-aspects of interrater agreement. Diagn Progn Res 2019/05/17 edn 2019;3:5.
43. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006;144(6):427–37.
44. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med 2013;158(4):280–6.
45. Bruls VEJ, Bastiaenen CHG, de Bie RA. Prognostic factors of complaints of arm, neck, and/or shoulder: a systematic review of prospective cohort studies. Pain 2015/02/07 edn 2015;156(5):765–88.
48. Gepstein R, Arinzon Z, Folman Y, Shabat S, Adunsky A. Lumbar spine surgery in Israeli Arabs and Jews: a comparative study with emphasis on pain perception. Isr Med Assoc J IMAJ 2007;9(6):443–7.
49. Kim HJ, Lee JI, Kang KT, Chang BS, Lee CK, Ruscheweyh R, et al. Influence of pain sensitivity on surgical outcomes after lumbar spine surgery in patients with lumbar spinal stenosis. Spine Phila Pa 1976 2014/11/11 edn 2015;40(3):193–200.
50. Kim HJ, Park JW, Chang BS, Lee CK, Yeom JS. The influence of catastrophising on treatment outcomes after surgery for lumbar spinal stenosis. Bone Jt J 2015;97-B(11):1546-1554.
51. Patel DV, Bawa MS, Haws BE, Khechen B, Block AM, Karmarkar SS, et al. PROMIS Physical Function for prediction of postoperative pain, narcotics consumption, and patient-reported outcomes following minimally invasive transforaminal lumbar interbody fusion. J Neurosurg Spine 2019;30(4):476–82.
52. Sigmundsson FG, Kang XP, Jonsson B, Stromqvist B. Prognostic factors in lumbar spinal stenosis surgery. Acta Orthop 2012/10/23 edn 2012;83(5):536–42.
53. Anderson PA, Schwaegler PE, Cizek D, Leverson G. Work status as a predictor of surgical outcome of discogenic low back pain. Spine 2006;31(21):2510–5.
54. Cushnie D, Thomas K, Jacobs WB, Cho RKH, Soroceanu A, Ahn H, et al. Effect of preoperative symptom duration on outcome in lumbar spinal stenosis: a Canadian Spine Outcomes and Research Network registry study. Spine J 2019;19(9):1470–7.
55. Ekman P, Moller H, Hedlund R. Predictive factors for the outcome of fusion in adult isthmic spondylolisthesis. Spine 2009;34(11):1204–10.
56. Støttrup CC, Andresen AK, Carreon L, Andersen MØ. Increasing reoperation rates and inferior outcome with prolonged symptom duration in lumbar disc herniation surgery — a prospective cohort study. Spine J 2019;19(9):1463–9.
57. Stromqvist F, Ahmad M, Hildingsson C, Jonsson B, Stromqvist B. Gender differences in lumbar disc herniation surgery. Acta Orthop 2008/10/08 edn 2008;79(5):643–9.

58. Muller M, Limacher A, Agten CA, Treichel F, Heini P, Seidel U, et al. Can quantitative sensory tests predict failed back surgery?: A prospective cohort study. Eur J Anaesthesiol 2019;36(9):695–704.
59. Hägg O, Fritzell P, Ekselius L, Nordwall A. Predictors of outcome in fusion surgery for chronic low back pain. A report from the Swedish Lumbar Spine Study. Eur Spine J 2003;12(1):22-33.
60. McGuire KJ, Khaleel MA, Rihn JA, Lurie JD, Zhao W, Weinstein JN. The effect of high obesity on outcomes of treatment for lumbar spinal conditions: subgroup analysis of the spine patient outcomes research trial. Spine Phila Pa 1976 2014/11/05 edn 2014;39(23):1975–80.
61. Radcliff KE, Rihn J, Hilibrand A, DiIorio T, Tosteson T, Lurie JD, et al. Does the duration of symptoms in patients with spinal stenosis and degenerative spondylolisthesis affect outcomes?: analysis of the Spine Outcomes Research Trial. Spine Phila Pa 1976 2011/09/14 edn 2011;36(25):2197–210.
62. Rihn JA, Radcliff K, Hilibrand AS, Anderson DT, Zhao W, Lurie J, et al. Does obesity affect outcomes of treatment for lumbar stenosis and degenerative spondylolisthesis? Analysis of the Spine Patient Outcomes Research Trial (SPORT). Spine Phila Pa 1976 2012/05/23 edn 2012;37(23):1933–46.
63. Schade V, Semmer N, Main CJ, Hora J, Boos N. The impact of clinical, morphological, psychosocial and work-related factors on the outcome of lumbar discectomy. Pain 1999;80(1–2):239–49.
64. Tampin B, Slater H, Jacques A, Lind CRP. Association of quantitative sensory testing parameters with clinical outcome in patients with lumbar radiculopathy undergoing microdiscectomy. Eur J Pain 2020;24(7):1377–92.
65. Watkins RG, O'Brien JP, Draugelis R, Jones D. Comparisons of preoperative and postoperative MMPI data in chronic back patients. Spine 1986;11(4):385–90.

2. Cited only in the supplement (additional reference numbers)

80. International Association for the Study of Pain. IASP Terminology. 2017. Available: https://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698#Centralsensitization. Accessed 5 Feb 2021.
81. Van Der Windt DA, Simons E, Riphagen II, Ammendolia C, Verhagen AP, Laslett M, Devillé W, Deyo RA, Bouter LM, de Vet HC. Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain. Cochrane Database Syst Rev 2010.
82. Fritsch EW, Heisel J, Rupp S. The Failed Back Surgery Syndrome: Reasons, Intraoperative Findings, and Long-term Results: A Report of 182 Operative Treatments. Spine 1996;21:626–633.
83. Khor S, Lavallee D, Cizik AM, Bellabarba C, Chapman JR, Howe CR, Lu D, Mohit AA, Oskouian RJ, Roh JR, Shonnard N, Dagal A, Flum DR. Development and Validation of a Prediction Model for Pain and Functional Outcomes After Lumbar Spine Surgery. JAMA Surg 2018;153:634–642.
84. Oosterhuis T, Smaardijk VR, Kuijer PPF, Langendam MW, Frings-Dresen MHW, Hoving JL. Systematic review of prognostic factors for work participation in patients with sciatica. Occup Env Med 2019;76:772–779.
85. Nicholas MK, Linton SJ, Watson PJ, Main CJ, Group "Decade of the Flags" Working. Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal. Phys Ther 2011;91:737–753.
86. 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;10:639–53.
87. Vlaeyen JWS, Linton SJ. Fear-avoidance model of chronic musculoskeletal pain: 12 years on. PAIN 2012;153:1144.
88. Modic MT, Pavlicek W, Weinstein MA, Boumphrey F, Ngo F, Hardy R, Duchesneau PM. Magnetic resonance imaging of intervertebral disk disease. Clinical and pulse sequence considerations. Radiology 1984;152:103–111.

89. Petersen KK, Vaegter HB, Stubhaug A, Wolff A, Scammell BE, Arendt-Nielsen L, Larsen DB. The predictive value of quantitative sensory testing: a systematic review on chronic postoperative pain and the analgesic effect of pharmacological therapies in patients with chronic pain. Pain 2021;162:31–44.
90. Husak AJ, Bair MJ. Chronic Pain and Sleep Disturbances: A Pragmatic Review of Their Relationships, Comorbidities, and Treatments. Pain Med 2020;21:1142–1152.
91. McBeth J, Lacey RJ, Wilkie R. Predictors of New-Onset Widespread Pain in Older Adults: Results From a Population-Based Prospective Cohort Study in the UK. Arthritis Rheumatol 2014;66:757–767.
92. Giusti EM, Lacerenza M, Manzoni GM, Castelnuovo G. Psychological and psychosocial predictors of chronic postsurgical pain: a systematic review and meta-analysis. Pain 2020/07/23 edn 2021;162(1):10–30.
93. Taylor RS, Desai MJ, Rigoard P, Taylor RJ. Predictors of pain relief following spinal cord stimulation in chronic back and leg pain and failed back surgery syndrome: a systematic review and meta-regression analysis. Pain Pr 2013/07/10 edn 2014;14(6):489–505.
94. OCEBM Levels of Evidence — Centre for Evidence-Based Medicine (CEBM), University of Oxford [Internet]. [cited 2021 Feb 23]. Available from: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence
95. Wupperman R, Davis R, Obremskey WT. Level of evidence in Spine compared to other orthopedic journals. Spine 2007;32(3):388–93.