**Systematic review and meta-analysis of predictors of return to work after spinal surgery for chronic low back and leg pain**

Halicka, Monikaa\*; Duarte, Ruib; Catherall, Sharonc; Maden, Michelleb; Coetsee, Michaelaa; Wilby, Martind; Brown, Christophera
aDepartment of Psychological Sciences, University of Liverpool, Liverpool, United Kingdom
bLiverpool Reviews & Implementation Group (LRiG), University of Liverpool, Liverpool, United Kingdom
cPublic Health Policy and Systems / LRiG, University of Liverpool, Liverpool, United Kingdom
dDepartment of Neurosurgery, The Walton Centre NHS Foundation Trust, Liverpool, United Kingdom

**\*Corresponding author**Email: mon.halicka@gmail.comPhone: +44 (0)1517941398Address: Department of Psychological Sciences, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool, L69 7ZA, United Kingdom

**Running title**Predictors of return to work after spinal surgery

**Disclosures**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.

# AbstractSpinal surgeries to treat chronic low back and/or leg pain (CLBP) have variable success rates, and despite the significant personal and socioeconomic implications, we lack usable prognostic indicators. This systematic review and meta-analysis evaluated the evidence for preoperative predictors of return to work (RTW) after spinal surgery for CLBP. We searched electronic databases and references (01/1984-03/2021), screened 2622 unique citations, and included 8 reports which involved adults with ≥3 months duration of CLBP undergoing first elective lumbar surgery with RTW outcomes assessed ≥3 months later. Five reports had low and 3 high risk of bias. We performed narrative synthesis and meta-analysis where possible and determined the quality of evidence (QoE). Individuals less likely to RTW after surgery were older (odds ratio [OR]=0.58; 95% confidence interval [CI]: 0.46 to 0.72), not working before surgery, had longer sick leave (OR=0.95; 95% CI: 0.93 to 0.97), higher physical workload, legal representation (OR=0.61; 95% CI: 0.53 to 0.71), psychiatric comorbidities and depression (moderate QoE), and longer CLBP duration and opioid use (low QoE), independent of potential confounders. We conclude that RTW after spinal surgery for CLBP likely depends on sociodemographic and affective psychological factors, and potentially also on symptom duration and opioid use.

# PerspectiveThis systematic review and meta-analysis synthesizes and evaluates existing evidence for preoperative predictors of return to work after spinal surgery for chronic low back pain. Demonstrated associations between return to work and sociodemographic, health-related, and psychological factors can inform clinical decision-making and guide further research.

# KeywordsChronic low back pain; spinal surgery; lumbar spine; predictors; return to work

# Introduction

Low back pain affects 40% of people at some point in their lives [31] and is the leading cause of years lived with disability in the world [44,52]. The personal and socioeconomic burden is particularly high for persistent or recurring low back pain, estimated to affect 60% of people a year after an initial acute episode [25]. When conservative treatment fails to reduce painful symptoms and improve function, surgical interventions can be considered to address the underlying spinal pathology [9,34]. The previous decade has seen more than a twofold increase in the number of lumbar spine surgeries performed in the UK and US [43,51]. While spinal surgery is one of the most invasive and expensive approaches to manage chronic low back pain (CLBP), its benefit is often suboptimal. A minimal clinically important reduction in pain intensity is reported only by 60% of patients undergoing first lumbar spine surgery [28,51,54,55] and long-term healthcare costs of CLBP that persists or recurs after surgery are 50% higher than those for patients without ongoing pain. Nearly 80% of the total cost of low back pain can be attributed to indirect costs associated with work absence and productivity loss [14].

Indeed, return to work (RTW) rates after lumbar spine surgery are highly variable (3-100%; [33]), partly due to patient heterogeneity. Identifying which factors can reliably predict the likelihood of RTW could inform development of clinical prediction models, facilitate managing patients’ expectations, which are strongly associated with work participation outcomes in CLBP more generally [26], and help to determine the most beneficial and cost-effective course of treatment based on individual socioeconomic, health-related, and psychological characteristics. For example, preoperative cognitive-behavioral therapy has been shown to improve outcomes of spinal fusion surgery [46].

Two previous systematic reviews suggested that patients who are older, female, have comorbidities and longer symptom duration, who are not working before surgery, have higher physical workload, occupational mental stress, passive pain coping, or depression are less likely to RTW after surgery for radiculopathy due to lumbar disc herniation [33,41]. These results were almost all based on single studies and could not be pooled for meta-analysis. Insufficient or conflicting evidence regarding prognosis of spinal surgery outcomes precludes development of clinical guidelines on relevant predictors [9,34]. Furthermore, the existing reviews were restricted to a specific population, while prognostic factors for RTW may be common across different spinal pathologies and surgical measures and thus provide stronger evidence for the examined associations. Importantly, chances of recovery and response to treatments decrease with longer duration of low back pain [13,15,25,40,49]. While this may affect the prognosis of RTW [48], so far there has been no evidence synthesis focusing on surgical candidates with CLBP. Therefore, we aimed to identify and evaluate preoperative predictors of RTW after spinal surgery for CLBP with or without radicular pain.

# Methods

This review was prospectively registered on PROSPERO (CRD42020180845) before commencing the screening stage. Our methodology and reporting followed the general principles for conducting reviews in health care outlined in the Centre for Reviews and Dissemination (CRD) guidance [10] and the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) [42]. The current article is one of the two planned review papers on the predictors of spinal surgery outcomes. The current paper focuses on predictors of return to work, and the second paper concerns predictors of patient-reported pain and disability outcomes. To summarize the findings regarding pain and disability outcomes, which will be reported elsewhere, there was moderate-quality evidence that greater preoperative sensory loss was associated with greater reduction in disability, and that symptom duration was unrelated to pain outcomes. There was low-quality evidence that younger age, higher education, and absence of spinal stenosis predicted better pain outcomes, but sociodemographic factors were unrelated to pain or disability outcomes, and comorbidities to disability outcomes. Finally, there was very low-quality evidence that lower preoperative pain, less comorbidities, lower pain catastrophizing, anxiety and depression were related to better pain outcomes, and lower job-related resignation and neuroticism to better disability outcomes, whereas disability and sensory testing were unrelated to pain reduction, and demographics, preoperative pain, and pain-related psychological factors were unrelated to change in disability.

## Search strategy

Electronic databases MEDLINE, EMBASE, PsycINFO, CINAHL, and Cochrane Central Register of Controlled trials (CENTRAL) were searched on 08/04/2020 and updated 29/03/2021, following a search strategy developed in collaboration with an information specialist (MM; for full search strategy, see **Supplementary Text S1**). Search results were exported to EndNote Library and de-duplicated. A manual search for additional primary studies was also conducted through the reference lists of relevant systematic reviews and included studies [7].

## Eligibility criteria

Eligibility criteria are outlined according to a modified PICOTS format for reviews of prognostic studies (Population, Index and Comparator prognostic factors, Outcomes, Timing, Setting) [72].

*Population* consisted of adults with CLBP, defined as pain lasting or recurring for ≥3 months [44], with or without lumbar radicular pain, defined as pain radiating to the leg due to nerve root compression [87]. Only patients undergoing primary lumbar or lumbosacral spine surgery, without history of previous spinal surgery, were eligible. We excluded spinal pathologies such as cancer / tumor, inflammatory disease, infection, or trauma, as well as spinal cord stimulator implantation, injections, radiofrequency, chemical interventions, and studies which investigated the impact of pre- or postoperative interventions.

*Index prognostic factors* includedvariables assessed prior to surgery and investigated for their potential ability to predict RTW after surgery. Radiographic, genetic, and any postoperative or intraoperative predictors were beyond the scope of the current review and were excluded, unless used as potential confounders. There is no standardized minimum set of *comparator prognostic factors* for spinal surgery outcomes*,* that is, ‘adjusted for’ factors when assessing the independent effect of certain index prognostic factors, therefore, we examined both unadjusted and adjusted prognostic effects, if reported.

Eligible studies reported *outcomes* including RTW, as an objective measure of functional recovery. The*timing*of outcome assessment was ≥3 months after surgery, with no upper limit. Study*settings*such as spinal surgery sites or registries of operated patients were eligible for inclusion.

Study *designs* had to allow investigation of associations between preoperative prognostic factors and postoperative RTW outcomes, and could include randomized or nonrandomized controlled studies, cohort, case-control, or registry-based studies, with prospective or retrospective designs. Case reports and case series were excluded as providing only low level of evidence in prognosis research [39,53].

*Publication* formats included original peer-reviewed studies published between 01/1984 (when diagnosis of spinal pathologies and surgical treatments could be informed by magnetic resonance imaging [37]) and 03/2021 in English language. Conference abstracts or unpublished research were not included.

## Study selection

Titles and abstracts of deduplicated records, and then full texts, were screened against the eligibility criteria by 2 independent reviewers (MH and RD) using a piloted form. Disagreements were resolved by consensus or seeking opinion from a third reviewer (MW) where necessary. If eligibility was unclear based on the abstract, the report was moved to full text screening. In case of uncertainty about eligibility based on the information available in the full text, supplement, and any related publications, study authors were contacted to request additional details. The selection process is illustrated in a PRISMA diagram [42] (**Figure 1**) and presents the primary reasons for exclusion recorded as the first category for which eligibility criteria were not met with certainty.

## Data extraction

A data extraction form was developed based on the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prognostic Factor Studies (CHARMS-PF) [45] and piloted by 2 independent reviewers (MH and SC; for the final template, see **Supplementary Table S1**). Each reviewer then extracted the data from half of the included reports and verified the accuracy of the data extracted by the other reviewer from the remaining reports. Any inconsistencies were resolved by discussion and consensus.

The following information was extracted from the included reports: study design, setting, method and time of participant recruitment and eligibility criteria, baseline participant characteristics, type of surgery, sample size estimation, number of included participants and participation rate, number, definition, method and time of measurement of index and comparator prognostic factors and outcomes, method of handling continuous factors and missing data, response rate, reasons for loss to follow-up and characteristics of lost participants, analysis methods, assumptions, methods and criteria for selecting predictors for and during multivariable modelling, adjusted and unadjusted effect estimates where available, signs of selective reporting, appropriateness of interpretation and discussion [45].

The desired common effect estimate for binary RTW outcomes was odds ratio (OR) with confidence interval (CI) or standard error (SE), however, depending on the analysis type, regression coefficients or mean differences were also extracted. Where the desired effect estimates were not provided, we calculated these from available data (e.g. 2x2 tables or means and standard deviations) or transformed the reported estimates using effect size calculators [18,35,36].

## Risk of bias assessment

Risk of bias (RoB) of included reports was assessed using the Quality in Prognosis Studies (QUIPS) tool [19,24,27] in 6 domains, each rated as being at high, moderate, or low RoB: study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis and reporting. RoB ratings were guided by prompting items suggested by Grooten et al. [19] and adapted to the current review question (for the final QUIPS template, see **Supplementary Table S2**). Two reviewers (MH and SC) independently assessed half of the included reports, and each checked the judgements of the other reviewer for agreement or domains with unclear RoB. Any inconsistencies were resolved by discussion and consensus. Finally, the overall ‘low’ or ‘high’ RoB ratings were assigned to each report, reflecting whether all QUIPS domains had low-moderate RoB (‘low’), or if ≥1 domain had high RoB (‘high’) [8]. Results of QUIPS assessment were considered during synthesis of the results and grading the quality of evidence.

## Data synthesis

Where there was sufficient and appropriate data for quantitative synthesis, we performed meta-analysis of the effects of index prognostic factors on RTW. Analyses were conducted in R software using *meta* package [5,23]. A minimum of 2 studies reporting data on the same predictor were required to pool the results. Precalculated prognostic effect estimates (ORs with SE or 95% CI) of included factors were pooled using random effects generic inverse variance. Random effects models were preferred as a degree of clinical and methodological heterogeneity was present among the included studies. The results of each meta-analysis were presented in a forest plot as pooled estimate of the average effect of the predictor with 95% CI and estimates of statistical heterogeneity (I2 and tau2). Tau2 was estimated using Paule-Mandel procedure as a recommended method when outcomes are binary and the number of pooled studies is small [4,50]. Substantial between-study heterogeneity was indicated by I2 >75% or 95% CI of tau2 not including 0 and p < .05 [29]. Meta-analyses were performed for the following candidate predictors deemed suitable for quantitative synthesis: age, gender, marital status, work status, duration of sick leave, worker’s compensation, legal representation, income, disability, and symptom duration. For a large proportion of these associations, only unadjusted effect estimates could be pooled. Due to small number of studies contributing to each meta-analysis, we did not conduct sensitivity analyses that would include only low RoB studies.

Meta-analyses were not conducted for other associations due to methodological heterogeneity or insufficient data. Specifically, it was not possible to quantitatively combine evidence from adjusted and unadjusted analyses, and with different definitions of predictors (e.g. timescales of opioid use and other analgesics). Insufficient results reporting (missing effect estimates and / or their precision), and only single studies contributing evidence for certain associations (e.g. personality traits / disorders, pain-related psychological factors), further prevented meta-analyses.

Since quantitative synthesis was only possible for some predictors, we included a tabular summary of all extracted adjusted and unadjusted associations with RTW, with a narrative synthesis of the findings. This synthesis presented the number of studies investigating each association of interest, the magnitude and direction of any reported effects, and assessed the consistency of available evidence between contributing studies. The evaluation of the findings also accounted for the RoB judgements at the study and outcome level.

## Grading of evidence

Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework [20] adapted for reviews of prognostic studies [32] guided the evaluation of the quality of evidence for associations between predictor categories and RTW. Two reviewers (MH and MC) collaboratively determined the overall quality of evidence as high, moderate, low, or very low, respectively reflecting high, moderate, limited, and very limited confidence that the true effect lies close to the observed estimate of that effect [20].

The starting quality of evidence depended on the phase of investigation, specifically, it was high for phase-3 and phase-2 studies (assessing prognostic pathways / mechanisms, and independent associations between hypothesized predictor and outcome, respectively), and moderate for phase-1 studies (exploring potential associations between prognostic factors and outcome) [32]. If any hypotheses-driven studies presented only unadjusted analyses or investigated broad classes of multiple predictors, they were classified as phase-1. Following the recommendations by Huguet et al. [32], the starting quality rating could be downgraded for severe study limitations (due to dominant evidence from high RoB studies or unadjusted analyses), clinically meaningful inconsistency in the effect estimates, indirectness (e.g. if it was not possible to verify chronic pain status of all included participants), imprecision (due to inadequate sample size or insufficient results reporting), and publication bias (if certain relationship was investigated in <4 studies); and upgraded if moderate (OR ≥ 2.5 or ≤ 0.4) or large (OR ≥ 4.25 or ≤ 0.24) effect size, or ‘dose’ effect (where higher levels of the predictor would lead to greater effect sizes) were present, resulting in an overall quality rating.

# Results

## ***Study selection***

We screened 2622 unique records and their flow through the selection process is illustrated in **Figure 1**. A list of excluded articles with reasons is provided in **Supplementary Text S2**. The most common reason for exclusion was ineligible population where patients without chronic pain or those who underwent previous spinal surgery were included. We were not able to confirm the chronic pain status of all participants in 3 studies, however, available average symptom duration data suggested that they were likely eligible [2,6,40]. This uncertainty was taken into account during the evaluation of the quality of evidence. Studies without any information about symptom duration were excluded. Reports that assessed surgical and nonsurgical cohorts were included if data specific to eligible surgical group could be extracted [22,38]. Overall, 8 eligible reports of 6 studies assessing predictors of RTW after surgery were included in the current review.

## Study characteristics

**Table 1** provides an overview of study characteristics. Five reports were based on prospective studies (2 single- [3,47] and 3 multi-center [6,21,22]). Since 2 of them analyzed the same participants from the Swedish Lumbar Spine Study, we considered them to reflect a single cohort [21,22]. Three reports included retrospective cohorts from the Ohio Bureau of Workers' Compensation [2,38,40]. They used the data from overlapping time periods, therefore, 2 reports that focused on fusion surgery were considered to correspond to a single cohort [2,38], while another report on single-level discectomy was considered to reflect a separate cohort [40]. Three of the included studies were based in the United States and the remaining ones in the Netherlands, Sweden, and Switzerland. The most common pathologies were disc herniation, degenerative disc disease and spondylosis, and surgical measures included fusion and discectomy. All participants were employed prior to surgery, except for Haag’s study (86% employed) [21,22], and unclear employment status in Anderson’s 2006 study as *not working* could also refer to being on sick leave [3]. Most studies defined RTW outcome as return to any (part- / full-time) work, 2 required RTW to be sustained for ≥6 months, and 1 considered only RTW in full capacity relative to before symptom onset. Follow-up duration ranged from 6 to 36 months, and RTW rates were between 26% and 90% across included studies.

## ***Risk of bias in included studies***

The reviewers made consistent RoB judgements on 83% of QUIPS domain ratings, with Cohen’s kappa = 0.70, 95% CI: 0.52 to 0.88, suggesting substantial agreement [1,12], before they reached final consensus. RoB ratings in each domain across all reports are illustrated in **Figure 2**. Overall, 5 reports were judged to have low, and 3 high RoB (where at least 1 QUIPS domain was rated as high) [8].

Most reports had low RoB in study participation, prognostic factor and outcome measurement, study confounding, and statistical analysis and reporting domains. Moderate RoB in study attrition domain in almost all included reports resulted from missing information about reasons for and / or characteristics of participants lost to follow-up, despite overall high attrition rates (min. 81% and >90% in most studies). RoB in 2 studies that retrospectively recruited only participants with complete follow-up [2,40] was also rated as moderate. Serious study limitations were identified in study participation, confounding, and statistical analysis and reporting domains in single studies. Specifically, high RoB in these domains was due to insufficient information regarding baseline characteristics, recruitment setting, and number of screened participants [3]; not adjusting for any confounders [22]; and missing information about statistical assumptions and partial results reporting [47].

## Results of syntheses

Predictors of RTW were examined in 6 unique patient cohorts (4616 participants in total). **Table 2** details the results for the associations analyzed in each included study. Where both adjusted and unadjusted effects were available, we only present the synthesis of independent associations below. Synthesis of unadjusted relationships can be found in **Supplementary Text S3**.

#### Sociodemographic predictors of RTW

Five studies (6 reports) investigated the associations between demographic and socioeconomic factors and RTW in 5 cohorts including a total of 4574 unique patients.

Older *age* was an independent negative predictor of RTW in 2 low RoB studies [2,21] but not in another high RoB study reporting adjusted analysis [3]. This observed inconsistency does not appear to be explained by sample characteristics, specific RTW criteria, duration of follow-up, or whether age was analyzed as a continuous or categorical factor; however, this discrepancy is unlikely to be clinically meaningful, and low RoB studies that did find a significant effect of age consistently reported its negative direction (including in unadjusted analyses). The pooled adjusted effect of age (defined as a categorical predictor, >48 or >50 years) supported a significant negative association (**Figure 3A**).

The independent effect of preoperative *work status* on RTW was examined in 1 phase-2 high RoB study, which found a large positive effect of working before surgery [3].

Three low RoB studies assessed the prognostic value of being on *sick leave* and / or its duration for RTW outcomes in adjusted analyses. Inability to continue working during the week before surgery negatively predicted RTW (small effect; [2]). Longer duration of sick leave was also a negative predictor of RTW in 3 studies (small effects; [6,21,38]), however, its independent effect in one study did not reach statistical significance [6]. Pooled adjusted OR of sick leave duration indicated a significant small negative effect on RTW (**Figure 3B**).

A relationship between having *legal representation* and RTW was assessed in 2 low RoB studies, both reporting small-moderate negative effects in adjusted analyses [2,38,40]. Pooled adjusted OR was consistent with a significant small negative effect of having legal representation on RTW (**Figure 3C**).

One low RoB study investigated the effect of *physical workload*, showing a small independent negative association between higher physical workload score and RTW [6].

An independent association between *income* and RTW was found in 2 low RoB studies, where participants with higher weekly wages and household income had higher odds of RTW after surgery (small effects; [38,40]). However, the pooled adjusted OR was not statistically significant and there was some indication of heterogeneity (**Figure 3D**).

We found no significant independent associations with RTW for other sociodemographic factors, that is, *gender* (1 high RoB study [3]), general *education* level (1 low RoB study [6]), or *workers’ compensation* status (1 high RoB study [3]). *Marital status* was only investigated in unadjusted analyses in 2 low RoB studies [21,38], with pooled OR indicating no significant effect (**Figure 3E**).

#### Health-related predictors of RTW

Six studies (7 reports) examined the associations between health- and symptom-related factors and RTW outcomes in 6 cohorts including a total of 4617 unique patients.

One low RoB study assessed whether type of *spinal pathology* affected RTW, showing a small independent effect of having spondylosis on lower odds of RTW [2]. Notably, several pathologies were considered in the adjusted analysis, yet only spondylosis was significant and reported.

Independent prognostic value of *symptom duration* for RTW was examined in 1 low RoB study. Longer time from injury to surgery was a significant negative predictor of RTW after surgery [40], and this association was supported by a pooled unadjusted OR from 3 other low RoB studies [6,21,38] (see Supplementary Text S3).

Two low RoB studies, including a phase-2 study, investigated the independent effect of *analgesics use* on RTW. Both found significant associations indicating that using (compared to not using) opioids before surgery [40], and using them for longer than (compared to less than) a year [2], was associated with lower odds of RTW (small effects).

No significant independent associations with RTW were found for other health-related factors, including preoperative *pain* (1 high RoB study [3]), *disability* (1 low and 2 high RoB studies [3,6,47]), sensory and motor *neurological signs* (1 low RoB study [6]), and *smoking* status (1 high RoB study [3]). Effects of acute *symptom onset* (1 low RoB study [6]), *comorbidities* (1 low RoB study [21]), and *BMI* (1 low RoB study [38]) were only investigated in unadjusted analyses, showing no significant associations with RTW after surgery.

#### Psychological predictors of RTW

Five studies (6 reports) assessed the relationships between psychological factors and RTW in 5 cohorts including 4510 unique patients in total.

One low RoB study assessed the association between *psychiatric comorbidity* and RTW. Participants with any psychiatric comorbidity (including affective disorders and schizophrenia) had lower odds of RTW after surgery in adjusted analysis (moderate effect size; [40]).

Two studies (1 low, 1 high RoB) [2,47] examined independent effects of *depression* on RTW. Participants with a clinical diagnosis of depression in one phase-2 study [2], and those with higher depression scores in another study [47], had lower odds of RTW relative to those without depression diagnosis or lower scores in adjusted analyses (moderate and small effect sizes, respectively).

The prognostic value of *pain coping* was assessed in one low RoB study, suggesting that passive pain coping was associated with reduced work capacity in adjusted analysis (small effect [6]).

The same low RoB study assessed the effect of *kinesiophobia* (fear of movement-related pain) on RTW, showing a small significant negative relationship with work capacity in adjusted analysis [6].

We found no significant independent associations between RTW and 2 work-related psychological factors, that is, *job satisfaction* (1 low RoB study [6]) and *occupational mental stress* (1 high RoB study [47]).

The prognostic value of several psychological factors was only examined in unadjusted analyses. A single high RoB study [47] reported significant negative relationships of *anxiety* and *job-related resignation*, and a positive relationship of *vitality,* with RTW after surgery, although effect estimates were not available. There were no significant associations between RTW and other psychological factors, including *pain behavior* (1 low RoB study [21]), *pain drawing* (1 high RoB study [22]), negative outcome *expectancies* (1 low RoB study [6]), and *personality* traits and disorders (1 low RoB study [21]; except for a small unadjusted negative effect of neuroticism).

## Quality of evidence

An overview of GRADE judgements of the quality of evidence is presented in **Figure 4**, and more detailed assessment is available in **Supplementary Table S3**.

There was *moderate*-quality evidence that demographic, socioeconomic, and affective psychological factors predict RTW. In particular, older age demonstrated independent negative prognostic value, gender, however, was unrelated to RTW outcomes. Regarding socioeconomic factors, participants who were employed before surgery were more likely to RTW even after accounting for potential confounders, whereas longer duration of sick leave, higher physical workload, and having legal representation showed independent negative associations with RTW. Income, workers’ compensation, and general education level did not predict RTW after adjusting for other factors. Among affective psychological factors, particularly depression and having any psychiatric comorbidity were found to be important independent negative predictors of RTW.

There was also *low*-quality evidence for independent prognostic value of symptom duration and analgesics use. Specifically, participants with longer duration of CLBP and those with opioids prescription and using opioids for longer were less likely to RTW after surgery.

The quality of evidence for the remaining associations was *very low*. Type of spinal pathology, in particular presence of spondylosis, and pain-related psychological factors such as passive coping and fear of movement, were identified as potential independent negative predictors of RTW. Finally, preoperative disability, pain intensity or pain in response to movement or touch, comorbidities, personality, and work-related psychological factors appeared unrelated to RTW outcomes.

The reasons for downgrading the quality of evidence can be summarized as follows. The data came largely from exploratory phase-1 studies, and only 3 of the examined relationships were supported by confirmatory phase-2 studies (analgesics use, depression, and work status; [2,3,40]). Lack of adjustment for any confounders further affected the confidence in some examined associations, especially in the psychological domain. Another issue that decreased the quality of evidence and limited the opportunities for its quantitative synthesis was imprecision related to inadequate sample size and insufficient results reporting, affecting half of the reviewed relationships. Limited number of studies for several candidate predictors presented potential publication bias.

# Discussion

We systematically reviewed the evidence for preoperative predictors of RTW after spinal surgery for CLBP, and performed narrative and quantitative synthesis of results where possible. Our main findings indicate that sociodemographic and affective psychological factors likely predict RTW. Symptom duration and opioid analgesic use also have potential prognostic value. The evidence for other health-related and psychological predictors is less certain.

## Sociodemographic predictors

We found moderate-quality evidence supported by meta-analysis that older age independently predicts decreased likelihood of RTW after spinal surgery, in line with previous systematic reviews [33,41].This seems particularly relevant in light of increasing prevalence of CLBP with age. Its burden will likely increase with aging population [30], and so may the number of people in need of spinal surgery.

The independent prognostic value of several socioeconomic work-related factors is supported by moderate-quality evidence. Patients working before surgery, with shorter duration of sick leave, lower physical workload, and without legal representation are more likely to RTW after surgery. These findings, with the effects of sick leave and legal representation confirmed in meta-analyses, reinforce the consistent evidence from previous reviews regarding RTW after surgery based largely on unadjusted associations [33,41], and from non-interventional studies on RTW in CLBP [48]. We further found that income may be less relevant to RTW, and that the previously reported effect of workers’ compensation claims on RTW [33] may lose its predictive ability after adjusting for work status. The value of socioeconomic predictors of RTW may depend on national healthcare and insurance systems, yet limited number of studies prevented sensitivity analysis to explore this further. Effects of legal representation, income, and work status were only assessed in US-based studies, however, association between sick leave and RTW was found across US, Swedish, and Dutch studies.

## Health-related predictors

Pertinent to our review question regarding CLBP, we found low-quality evidence for an independent effect of longer symptom duration on reduced RTW, further supported by their pooled unadjusted association. These findings are in agreement with a previous review of RTW rates after spinal surgery [33], and with strong evidence from non-surgical CLBP studies that delay in referral for intervention has adverse effects on RTW [48]. Preoperative opioid prescription and prolonged opioid use also independently predicted reduced RTW in the current review (low-quality evidence). Opioids are commonly prescribed for moderate and severe pain that could not be managed with other treatments, therefore these patients likely represent more severe cases [16]. Very low-quality evidence suggests that spinal pathology, particularly presence of spondylosis, may be a negative independent predictor of RTW. Patients with spondylolysis are often older and present with more persistent symptoms. Overall, these prognostic effects, in line with the duration of sick leave, appear to suggest that chances of RTW decrease with increasing duration and severity of CLBP. However, preoperative disability (seemingly closely related to functional recovery indexed by RTW) only predicted this outcome in unadjusted meta-analysis, but not after adjusting for socioeconomic and psychological factors (very low-quality evidence).

## Psychological predictors

Recommendations regarding the prognostic value of psychological distress for spinal surgery outcomes are scarce and vary across countries. For instance, in the case of disc herniation with radiculopathy, the UK National Institute for Health and Care Excellence advises not to use such information during patient selection for surgery, whereas the North American Spine Society suggests that patients should be assessed for signs of psychological distress, such as somatization and depression, based on fair evidence that these signs predict worse outcomes [9,34]. Our review, considering a broader range of degenerative spine diseases, found no evidence for pain behavior or non-organic signs, but passive pain coping and fear of movement may be related to RTW after surgery (very low-quality evidence). Importantly, we found moderate-quality evidence that patients who are depressed or have a psychiatric comorbidity are less likely to RTW, even after controlling for potential confounders. Although the data was not suitable for meta-analysis, this evidence was supported by moderate effect sizes from large studies, including a confirmatory investigation. Unadjusted effects of anxiety and low vitality were consistent with the conclusion that negative affective factors likely predict reduced RTW.

## Strengths and limitations

This review provides a comprehensive evaluation of 33 candidate predictors for RTW, objectively reflecting functional recovery after surgery. We expand upon the previous literature [33,41] by including a broader range of degenerative diseases of the lumbar spine, while specifically focusing on patients with chronic pain. Identifying a larger set of relevant studies allowed to pool the effects of some associations in meta-analysis. Furthermore, our robust quality assessments provide a transparent overview of the certainty in available prognostic evidence in this filed, highlight prevalent methodological issues, and signpost directions for further research.

Low quality of evidence was a concern for predictors other than sociodemographic and affective psychological factors. Specific shortcomings of the available evidence are outlined in the *Risk of bias* and *Quality of evidence* results sections. The current review also has some limitations. First, several examined associations included indirect evidence, where it was not possible to verify chronic pain status of all participants [2,6,40], thus these samples may not accurately reflect the review question. Nonetheless, available data strongly suggested their eligibility and any uncertainty was reflected in the quality assessments. Second, in 2 included studies, not all patients were working before surgery[3,21,22]*.* While this may seem suboptimal considering RTW outcome, it allowed to examine the prognostic effect of preoperative work status. Third, definitions of RTW were not consistent across all studies, with some specifying sustained RTW, and others work capacity. Although all these definitions were relevant to our review question, they added a degree of study heterogeneity.However, outcome definition was not identified as a potential source of inconsistencies in the results*.* Ideally, the impact of the above-mentioned limitations should be examined in sensitivity analyses, yet these were not possible due to the small number of studies.This also limits our confidence in the precision of pooled estimates, based only on 2-3 studies for each association examined in meta-analysis. It also affected the feasibility of quantitative synthesis for other candidate predictors. Moreover, moderate or large effect sizes contributed only to prognostic effects of socioeconomic and affective factors, while the magnitude of all pooled estimates was small. Finally, the current article is part of a broader review including predictors of patient-reported pain and disability outcomes. While the decision to report them in a separate article was not made *a priori*, it allows to provide more in-depth assessment of these distinct outcomes.

## Implications

The identified limitations of the existing evidence suggest specific directions for further research. Issues preventing meta-analysis could be addressed at the levels of study design and results reporting, where consensus on consistent measures and definitions of the same predictors and outcomes would reduce heterogeneity [11,17]; monitoring and reporting attrition could ensure representativeness of the studied samples; and transparent reporting of both positive and negative results with precision estimates would allow quantitative synthesis and increase certainty in presented evidence. This review further highlights the importance of controlling for potential alternative explanations in prognosis research. For instance, possible effects of education level, workers’ compensation, disability, and work-related psychological factors were found in unadjusted analyses, but could not be replicated after adjusting for other factors. As there is no recommended set of relevant confounders, we suggest that age, socioeconomic, and affective factors should be adjusted for in future prognosis research. Additional high-quality confirmatory studies should verify the independent prognostic value of symptom duration, analgesics use, pain-related psychological factors, and spinal pathology, which is currently supported by low or very low-quality evidence. Despite overall small effect sizes arising from the current syntheses, combining several prognostic factors could increase the accuracy of outcome prediction, thus the identified likely predictors of RTW after spinal surgery should be considered in the development of clinical prediction models.

There are several practical implications of the presented findings. For instance, the negative effects of socioeconomic factors and symptom duration on RTW suggest that patients might benefit from being operated earlier in the course of CLBP, or working as long as possible until surgery but perhaps with reduced workload. Furthermore, patients with CLBP with signs of depression or maladaptive pain coping may benefit from preoperative cognitive-behavioral therapy [46]. Ideally, multidisciplinary approaches to pain management should be sought as an alternative to opioid pain relief, which, if necessary, should only be prescribed for short periods.

## Conclusions

The likelihood of RTW after spinal surgery for CLBP appears to depend on patients’ demographics, socioeconomic situation, medical history, and affective psychological characteristics. We found likely negative prognostic value of older age, longer sick leave, having legal representation, higher physical workload, not working before surgery, having psychiatric comorbidities and depression for RTW. Longer symptom duration and opioid use also potentially predict reduced RTW, whereas the prognostic value of other preoperative factors is less certain. The current level of evidence may not be sufficient for the development of clinical guidelines regarding prognosis, and more high-quality prospective data would increase confidence in the above associations [53]. However, the identified predictors could inform the design of future confirmatory studies and clinical prediction models, help to estimate the likelihood of functional recovery and choose the best course of treatment at the right time, to reach individual patient goals and maximize the benefit from surgery.

# Acknowledgements

We thank Dr Sarah Nevitt for her advice on the systematic review protocol.

# Supplemental Digital Content

# Supplementary Text S1: Search strategies for electronic databases; Supplementary Table S1: Data extraction form template; Supplementary Table S2: Risk of bias assessment form template; Supplementary Text S2: Excluded full text articles with reasons; Supplementary Text S3: Results of syntheses of unadjusted associations with return to work; Supplementary Table S3: GRADE quality of evidence assessment.

# References

[1] Altman DG. Practical statistics for medical research. CRC press, 1990.

[2] Anderson JT, Haas AR, Percy R, Woods ST, Ahn UM, Ahn NU. Clinical depression is a strong predictor of poor lumbar fusion outcomes among workers’ compensation subjects. Spine 2015;40:748–56.

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

[4] Bakbergenuly I, Hoaglin DC, Kulinskaya E. Methods for estimating between-study variance and overall effect in meta-analysis of odds ratios. Res Synth Methods 2020;11:426–442.

[5] Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. Evid Based Ment Health 2019;22:153–160.

[6] den Boer JJ, Oostendorp RA, Beems T, Munneke M, Evers AW. Reduced work capacity after lumbar disc surgery: the role of cognitive-behavioral and work-related risk factors. Pain 2006;126:72–8.

[7] Bramer WM. Reference checking for systematic reviews using Endnote. J Med Libr Assoc JMLA 2018;106:542.

[8] 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;156:765–788.

[9] de Campos TF. Low back pain and sciatica in over 16s: assessment and management NICE Guideline [NG59]. J Physiother 2017;63:120.

[10] Centre for Reviews and Dissemination. Systematic reviews: CRD’s guidance for undertaking reviews in healthcare. University of York: CRD, 2009.

[11] Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LO, Foster NE, Grotle M, Koes BW, Kovacs FM. Core outcome measurement instruments for clinical trials in nonspecific low back pain. Pain 2018;159:481.

[12] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.

[13] 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:E613–E624.

[14] Dagenais S, Caro J, Haldeman S. A systematic review of low back pain cost of illness studies in the United States and internationally. Spine J 2008;8:8–20.

[15] Dhondt E, Van Oosterwijck J, Cagnie B, Adnan R, Schouppe S, Van Akeleyen J, Logghe T, Danneels L. Predicting treatment adherence and outcome to outpatient multimodal rehabilitation in chronic low back pain. J Back Musculoskelet Rehabil 2020;33:277–293.

[16] Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain—United States, 2016. JAMA 2016;315:1624–1645.

[17] Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, Kerns RD, Stucki G, Allen RR, Bellamy N. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9–19.

[18] Effect Size Calculator. n.d. Available: https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php. Accessed 23 Feb 2021.

[19] Grooten WJA, Tseli E, Ang BO, Boersma K, Stalnacke BM, Gerdle B, Enthoven P. 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;3:5.

[20] 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;64:380–382.

[21] 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;12:22‐33.

[22] Hägg O, Fritzell P, Hedlund R, Möller H, Ekselius L, Nordwall A. Pain-drawing does not predict the outcome of fusion surgery for chronic low-back pain: a report from the Swedish Lumbar Spine Study. Eur Spine J 2003;12:2‐11.

[23] Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing Meta-Analysis With R: A Hands-On Guide. 1st ed. Boca Raton, FL and London: Chapman & Hall/CRC Press, 2021 Available: https://www.routledge.com/Doing-Meta-Analysis-with-R-A-Hands-On-Guide/Harrer-Cuijpers-Furukawa-Ebert/p/book/9780367610074.

[24] Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006;144:427–437.

[25] Hayden JA, Dunn KM, van der Windt DA, Shaw WS. What is the prognosis of back pain? Best Pract Res Clin Rheumatol 2010;24:167–179.

[26] Hayden JA, Wilson MN, Riley RD, Iles R, Pincus T, Ogilvie R. Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor review. Cochrane Database Syst Rev 2019;2019. doi:10.1002/14651858.CD011284.pub2.

[27] Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med 2013;158:280–286.

[28] Hegarty D, Shorten G. Multivariate prognostic modeling of persistent pain following lumbar discectomy. Pain Physician 2012;15:421–34.

[29] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–1558.

[30] Hoy D. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014;10.

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

[32] Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, Wozney L. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. Syst Rev 2013;2:1–12.

[33] Huysmans E, Goudman L, Van Belleghem G, De Jaeger M, Moens M, Nijs J, Ickmans K, Buyl R, Vanroelen C, Putman K. Return to work following surgery for lumbar radiculopathy: a systematic review. Spine J 2018;18:1694–1714.

[34] Kreiner DS, Hwang S, Easa J, Resnick DK, Baisden J, Bess S, Cho CH, DePalma MJ, Dougherty P, Fernand R, Ghiselli G, Hanna AS, Lamer T, Lisi AJ, Mazanec DJ, Meagher RJ, Nucci RC, Patel RD, Sembrano JN, Sharma AK, Summers JT, Taleghani CK, Tontz WL, Toton JF. Diagnosis and Treatment of Lumbar Disc Herniation with Radiculopathy. North American Spine Society, 2012.

[35] Lenhard W, Lenhard A. Computation of Effect Sizes. 2017. doi:10.13140/RG.2.2.17823.92329.

[36] Lüdecke D. esc: Effect Size Computation for Meta Analysis. Zenodo, 2018 doi:10.5281/zenodo.1249218.

[37] 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.

[38] Nguyen TH, Randolph DC, Talmage J, Succop P, Travis R. Long-term outcomes of lumbar fusion among workers’ compensation subjects: a historical cohort study. Spine 2011;36:320–31.

[39] OCEBM Levels of Evidence — Centre for Evidence-Based Medicine (CEBM), University of Oxford. n.d. Available: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence. Accessed 23 Feb 2021.

[40] O’Donnell JA, Anderson JT, Haas AR, Percy R, Woods ST, Ahn UM, Ahn NU. Preoperative Opioid Use is a Predictor of Poor Return to Work in Workers’ Compensation Patients After Lumbar Diskectomy. Spine 03622436 2018;43:594–602.

[41] 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.

[42] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.

[43] Rajaee SS, Bae HW, Kanim LEA, Delamarter RB. Spinal fusion in the United States: analysis of trends from 1998 to 2008. Spine 2012;37:67–76.

[44] Rice AS, Smith BH, Blyth FM. Pain and the global burden of disease. Pain 2016;157:791–796.

[45] Riley RD, Moons KGM, Snell KIE, Ensor J, Hooft L, Altman DG, Hayden J, Collins GS, Debray TPA. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ 2019;364:k4597.

[46] 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;41:18–25.

[47] 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;80:239–49.

[48] Steenstra IA, Munhall C, Irvin E, Oranye N, Passmore S, Van Eerd D, Mahood Q, Hogg-Johnson S. Systematic Review of Prognostic Factors for Return to Work in Workers with Sub Acute and Chronic Low Back Pain. J Occup Rehabil 2017;27:369–381.

[49] 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;154:1237–1244.

[50] Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, Kuss O, Higgins JP, Langan D, Salanti G. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res Synth Methods 2016;7:55–79.

[51] Weir S, Samnaliev M, Kuo T-C, Choitir CN, Tierney TS, Cumming D, Bruce J, Manca A, Taylor RS, Eldabe S. 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:e017585.

[52] Wu A, March L, Zheng X, Huang J, Wang X, Zhao J, Blyth FM, Smith E, Buchbinder R, Hoy D. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. Ann Transl Med 2020;8:299.

[53] Wupperman R, Davis R, Obremskey WT. Level of evidence in Spine compared to other orthopedic journals. Spine 2007;32:388–393.

[54] Zweig T, Enke J, Mannion AF, Sobottke R, Melloh M, Freeman BJ, Aghayev E, Spine Tango C. 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 2017;26:488–500.

[55] 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;12:275–275.

# Figure legends

**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram [42].

**Figure 2.** Risk of bias judgements based on Quality in Prognosis Studies (QUIPS) [27]. L, low; M, moderate; H, high risk of bias. Overall assessment of RoB: Low = all domains low or moderate; High = one or more domains high [8].

**Figure 3.** Forest plots of pooled effects of (A) age, (B) duration of sick leave, (C) legal representation, (D) income, and (E) marital status on return to work after surgery. If standard error or confidence interval of effect estimate was missing, that effect was not included in meta-analysis, but its magnitude is presented alongside other eligible studies (B). adj., adjusted; CI, confidence interval; logOR, log-odds ratio; OR, odds ratio; SE, standard error; unadj., unadjusted effect.

**Figure 4.** Overall quality of evidence for the reviewed associations according to Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework [20,32].