Malta Journal of Health Sciences
https://doi.org/10.14614/PROGNOSTIC/12/25

Prognostic models of non–surgical treatment outcomes for lumbar–related leg pain: a scoping review of systematic reviews.

Mr Emanuel Schembri
Department of Physiotherapy, Faculty of Health Sciences, University of Malta.
Dr Karl Spiteri
Department of Physiotherapy, St Vincent de Paul, Malta.
Dr Laura Jayne Bonnett
Department of Health Data Science, University of Liverpool.
Dr John Xerri de Caro
Department of Physiotherapy, Faculty of Health Sciences, University of Malta.

Abstract

Prognosis identifies a relationship between future outcomes and a given health state. Prognostic factors help estimate the likelihood of a particular outcome, regardless of any specific intervention. The prognosis of persons treated for lumbar–related leg pain (LRLP) varies considerably. To the authors' knowledge, no systematic review (SR) has evaluated any formal combination of multiple predictors (prognostic models) for non–surgical interventions in this population. This scoping review aimed to: (a) determine if SRs specifically evaluating prognostic models of non–surgical interventions for LRLP exist, and (b) identify prognostic models addressing non–surgical interventions in LRLP within the SRs. A systematic search of PubMed, Embase, CINAHL, and Epistemonikos was conducted in October 2024, including SRs published from the year 2000 onwards. References were hand–searched, and forward citation searches were performed. The search identified 9,398 records after deduplication. Following screening, the full texts of 18 SRs were evaluated against the eligibility criteria. The lack of SRs in this area highlights a critical gap in understanding the prognostic models for non–surgical interventions in LRLP. Three reviews reported two primary studies that had derived and/or evaluated prognostic models in cohorts that met the threshold set for this scoping review, with 80% of participants experiencing LRLP. This scoping review highlights a gap in the SR literature evaluating specific prognostic models for non–surgical interventions in individuals with LRLP even though various prognostic models for LRLP exist. This underlines the need for a dedicated SR to consolidate evidence and guide clinical practice in the non–surgical management of LRLP.

Keywords: Sciatica; Radiculopathy; Models, Statistical; Predictive model; Scoping review.

Correspondence to Emanuel Schembri
esche05@um.edu.mt
Received: 19.12.2024
Revised: 19.2.2025
Accepted: 8.3.2025
Published: 30.6.2025

1. Introduction

Acute episodes of referred leg pain that originate from the lumbar region normally resolve within a few weeks as a consequence of the natural course of healing. However, outcomes appear to be poorer than those episodes of low back pain presenting without leg pain (Konstantinou *et al.*, 2018). Studies have reported that following the conservative management of persons with lumbar radicular pain, only 35.7% had a good outcome at 12 months (Azharuddin *et al.*, 2022) and 42% needed surgery after one year (Boden *et al.*, 2018). 70% of the patients referred with symptomatic lumbar spinal stenosis reportedly remained the same or worsened or

required surgical treatment after three years (Matsudaira *et al.*, 2016).

Episodes that present as referred pain along the leg are often labelled as sciatica, giving the impression that sciatica is more than a symptom. Definitions for sciatica vary with distinctions being reported, for example, between self–reported and clinically assessed symptoms, which leads to a poor interpretation of clinical outcomes (Konstantinou & Dunn, 2008). Sciatica is often confused with other sources of somatic referred leg pain. This confusion contributes to wide discrepancies in prevalence rates (1.6 to 43%) and complicates effective patient management (Konstantinou & Dunn, 2008). Consequently, the International Association for the Study of Pain (IASP) NeuPSIG recommends avoiding the term 'sciatica' in favour of 'spine–related leg pain' as a more precise umbrella term that addresses both somatic referred pain and radicular pain with or without radiculopathy (Schmid *et al.*, 2023). In this present article, we adopt the term lumbar–related leg pain (LRLP) to specifically focus on the lumbar spine.

Prognosis research investigates the relations between future outcomes (endpoints) among people with a given health state (start point) to improve health. The PROGnosis RESearch Strategy (PROGRESS) series is a framework of four distinct but inter–related prognosis research themes being: overall prognosis, prognostic factors, prognostic models and stratified medicine. The PROGRESS series aims to explain how each of these four themes provides evidence that can be used at multiple (translational) pathways toward improving clinical outcomes—from the discovery of new interventions, through to their evaluation and implementation in the clinical management of individual patients, and to examine the impact of interventions and healthcare policies on patient outcomes (Hemingway *et al.*, 2013; Riley *et al.*, 2013; Steyerberg *et al.*, 2013; Hingorani *et al.*, 2013).

Multivariable prediction models are categorised into diagnostic and prognostic models based on their purpose and temporal focus. Diagnostic models combine multiple predictors, often diagnostic test results, to estimate the probability of a disease or condition being present or absent at the time of prediction (Moons *et al.*, 2015). Prognostic models integrate multiple predictors to estimate the probability of a specific outcome or event occurring in the future. These models apply to individuals at risk of that outcome, whether healthy or ill. They can forecast events such as recurrence, complications, or mortality within a defined timeframe, including in individuals without a diagnosed disease (D'Agostino *et al.*, 2008). The term prognostic is used here in a broader sense, encompassing predicting future outcomes in at–risk populations rather than being limited to forecasting the progression of patients with a specific disease, regardless of treatment. The key distinction between diagnostic and prognostic models is time: diagnostic models assess the present and are developed through a cross–sectional research design. Prognostic models forecast future outcomes and are usually longitudinal (Moons *et al.*, 2015).

The STarT Back tool (SBT) is a clinical prediction tool in the form of a questionnaire that assesses the likelihood of disability and chronicity in six months' time in a predominantly acute cohort, and it consists of nine items that screen for physical and psychological predictors of persistent disabling low back pain (LBP). The binary scores are added, and the total score stratifies responses into three pre–defined subgroups: low–risk, medium–risk, and high–risk of persistent disabling LBP. Each risk subgroup has a pre–defined treatment matched to it. This is recommended for the stratification of care pathways in the UK primary care setting (Hill *et al.*, 2011). The SBT was a considerable step in the personalised care approach for patients with LBP.

The NICE guidelines on LBP and sciatica (NICE, 2020a) included a reference to the adoption of the SBT with a disclaimer stating that the quality and usability of the tool had not yet been judged (NICE, 2020b). The inclusion of exclusively modifiable prognostic factors in the SBT does not provide a 360° approach to making accurate predictions since non–modifiable factors also have a predictive role (Parreira *et al.*, 2018). The value of a screening instrument is also directly related to setting–specific conditions and optimal in the cohorts for which it was developed (Karran *et al.*, 2017a). Studies have reported that the final outcomes for managing LBP following the SBT approach were superior in the UK primary care settings in which the SBT was developed and validated, compared to other settings that adopted the SBT approach, such as physiotherapy and chiropractic settings, as well as for secondary levels of care in which the SBT was not validated (Morsø *et al.*, 2014; Karran *et al.*, 2017b). The predictive value of the SBT also did not prove advantageous in persons suffering from chronic low back pain (Kendell *et al.*, 2018), most likely because the tool was developed for a cohort largely composed of

acute cases of LBP (72–74% of the participants) (Hill *et al.*, 2011).

To the authors' knowledge, no systematic review (SR) evaluated predictive models of non–surgical interventions specifically for individuals with LRLP. The primary aim of this scoping review was to: (a) determine if SRs specifically evaluating prognostic models of non–surgical interventions for LRLP exist, and (b) to identify prognostic models addressing non–surgical interventions in LRLP within the SRs.

Therefore, this scoping review aims to answer the primary research question: Are there SRs that evaluate prognostic models of non–surgical interventions specifically related to LRLP? The secondary research question was: Are there prognostic models addressing non–surgical interventions for individuals with LRLP?

2. Methods

2.1. Reason for conducting a scoping review

This scoping review aimed to determine if the literature already reported a SR on the evaluation of prognostic models of treatment outcomes in the non–surgical management of LRLP. The broader research aims, and the exploratory nature of the envisioned literature review necessitated a methodology that would map and summarise the evidence and extent of knowledge to better inform further research (Peters *et al.*, 2020). Ensuring that a SR has not already been conducted on the topic helps prevent unnecessary duplication of efforts. For this reason, a scoping review was considered particularly suitable for exploring this aim since such reviews are exploratory projects that systematically map the literature available on a topic, identifying the key concepts, theories, sources of evidence, and gaps in the research (Canadian Institutes of Health Research).

2.2. Eligibility criteria

This scoping review included SRs (with or without meta–analyses) that were published in the English language, which evaluated a population composed of adults over 18 years reporting LBP, specifically requiring at least 80% of the study cohort to have LRLP, and which may include sciatica, neurogenic claudication, radicular pain, or a combination thereof. Studies were excluded if less than 80% of participants had LRLP or if pain originated from other causes, such as tumours or fractures (Table 1). The 80% threshold was designed in a pragmatic manner to limit the contamination of the sample with other types of LBP.

The reviews had to include studies that evaluated prognostic models of non–surgical outcomes, including medication, physiotherapy, epidural steroid injections, and multidisciplinary rehabilitation. Studies addressing models predicting the outcomes following surgical interventions or models predicting pain chronicity or disability unrelated to a specific treatment were excluded.

This scoping review aimed to identify prognostic models that adhere to the definition provided within TRIPOD (Moons *et al.*, 2015, p. 1), where a prognostic model is defined as the "mathematical equation that estimates the probability of having a disease or condition in the present (diagnostic prediction model) or the probability of developing a particular disease or outcome in the future (prognostic prediction model)". Therefore, for continuous outcomes, a prognostic model predicts an individualised expected (mean) outcome value by a particular time point. For binary or time–to–event outcomes, a prognostic model predicts an individual outcome risk (probability) by a particular time point (or time points) (Steyerberg *et al.*, 2013).

Due to the studies' methodological heterogeneity, especially in the early years of prognostic medicine, clinical prediction rules (CPRs) and screening tools for stratifying patients into meaningful prognostic subgroups will also be eligible in this review (Fu *et al.*, 2024). CPRs are simple statistical prediction tools designed to be used with individual patients. They comprise a small number of clinical variables that have been identified to be independently predictive of a given diagnosis, outcome, or treatment effect. These can be both prognostic and prescriptive CPRs, where the former consists of prognostic variables that inform predictions of future outcomes while the latter is a special type of prognostic CPRs since they inform predictions regarding the relative treatment effect a patient may experience from an intervention (Foster *et al.*, 2013). The variables included in a prescriptive CPR are treatment effect modifiers (Kraemer, Frank & Kupfer, 2006); thus, they function to inform clinical decisions regarding treatment selection (Cook, 2008).

Any identified prognostic model's predictive ability, including accuracy, calibration, discrimination, net benefit, and R^2 values were extracted (PROGRESS 3). If any of the prognostic models were evaluated within a

Table 1. PIOS criteria for inclusion in the review.

	Inclusion	Exclusion
Population	Adults (>18 years) with low back pain† disorders, having at least 80% of the participants with lumbar–related leg pain (including but not limited to sciatica or neurogenic claudication or radicular pain or a mixture of the three).	Less than 80% of the cohort having lumbar–related leg pain. Back pain or lumbar–related leg pain due to other conditions, e.g., tumour or fracture.
Intervention	Prognostic models of treatment outcomes for lumbar–related leg pain following non–surgical treatments (for example, but not limited to, medications, physiotherapy, epidural steroid injections, and multidisciplinary rehabilitation).	Models predicting outcomes following spinal surgeries or other conditions. Models/factors predicting disability or pain chronicity but not associated with any formal treatment/ intervention.
Outcome	Predictive abilities of the prognostic models, including calibration, discrimination, net benefit, and R^2 values. Measures of benefit in case a model was evaluated within an RCT (PROGRESS level 4).	The study solely focuses on treatment outcomes without evaluating the predictive ability of the models.
Study design	**Research question 1:** Systematic reviews with or without meta–analyses of prognostic model studies. **Research question 2:** Primary studies included within the systematic reviews screened by full text, adopting either an observational (cohort, registry or retrospective studies) or RCT design that also fulfils the first three aspects of the PIO criteria.	All other study designs, reviews that do not have an English language version.

†Low back pain was included since studies frequently evaluate back pain and lumbar–related leg pain together. However, to be eligible for inclusion, a prognostic model developed or validated in a mixed population must have at least 80% of the cohort composed of persons with lumbar–related leg pain.

randomised controlled trial (RCT) (PROGRESS 4), the outcomes were briefly summarised. Studies without predictive assessments were excluded.

2.3. Critical appraisal

Scoping reviews usually do not necessitate critically appraising the included articles (Peters *et al.*, 2020). However, this review incorporated an appraisal process for specific cases. If a SR examining prognostic models of treatment outcomes for LRLP was identified, it was critically appraised using the AMSTAR–II tool (Shea *et al.*, 2017). Similarly, if a prognostic model of treatment outcomes for persons with LRLP was identified, it was appraised using the PROBAST tool (Wolff *et al.*, 2019). Two reviewers (ES, KS) independently conducted the critical appraisal, and any discrepancies were resolved through consensus.

2.4. Information sources and search strategy

The scoping review was carried out following the PRISMA for scoping review guidelines (PRISMA–ScR) (Tricco *et al.*, 2018). The search strategy was developed by ES and checked by an academic librarian at the University using the PRESS checklist (McGowan *et al.*, 2016). An electronic literature search was conducted on the 7th

of October 2024 within PubMed, Embase (Elsevier), CINAHL (EBSCO), and Epistemonikos (Goossen *et al.*, 2020). References of the included reviews were hand–searched, and forward citation searching using the software CitationChaser (Haddaway, Grainger & Gray, 2022) was conducted. Due to the inconsistent reporting standards, reliability, and data adequacy often associated with grey literature, this was not searched.

The focus of the search was on LRLP. However, terms related to LBP were included in the search since both conditions are frequently studied together. Three main facets were incorporated into the search strategy: LRLP, LBP, and a sensitive search filter for prognostic/prediction models (Appendix A).

A combination of free–text keywords, their synonyms and, where appropriate, word truncation was employed in the search strategy. Furthermore, any relevant medical subject headings (MeSH terms) or Emtree terms were incorporated into the search. The first two facets were combined using the Boolean operator "OR" and subsequently joined with the search filter using the Boolean operator "AND". The year of publication was limited to the period 2000–2024, reflecting the publication timeline of the Orebro Musculoskeletal Questionnaire, a prognostic tool for back pain disorders, which was published in 2003 (Linton & Boersma, 2003).

2.5. Search Filters

To retrieve prognostic model studies within the three databases, validated search filters that were specifically designed to retrieve prognostic studies within CINAHL (Walker–Dilks, Wilczynski & Haynes, 2008), Embase (Holland, Wilczynski & Haynes, 2005) and PubMed (Geersing *et al.*, 2012) were used within the respective search strings (Appendix A). PubMed was preferred over other platforms, such as Medline, due to the unique aspects of the PubMed search engine. The Geersing *et al.*, (2012) search filter yielded higher sensitivity values [(0.97; 95% CI 0.83 to 0.99) (0.94; 95% CI 0.74 to 0.99)], and the lowest number needed to read (NNR = 68 to 125) values to justify it.

2.6. Screening and data charting

Two reviewers (ES, KS) collaboratively designed a data–charting form to identify the variables for extraction. The same reviewers independently screened the retrieved titles against the eligibility criteria. Liaison with a third reviewer (JXDC) was sought in case of disagreements. The same reviewers independently extracted data from the identified articles using a standardised data extraction sheet. In case a SR identified a model that was specifically developed, validated or tested in a cohort of persons with LRLP, the authors also sought related papers on its development, validation and/or updating to aid data extraction and critical appraisal. In such cases, the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) checklist (Moons *et al.*, 2014) was used to extract data. In cases of disagreement or conflict during the data extraction phase, a third reviewer (JXDC) was consulted to achieve consensus through discussion.

3. Results

3.1. Selection of sources of evidence

Epistemonikos (n=907), CINAHL (EBSCO) (n=1,522), PubMed (n=1,720), and Embase (Elsevier) (n=7,762) were searched on October 7, 2024, for a total of 11,911 records. After deduplication, the number of records was 9,398, which were screened by their title and abstract (Figure 1). The full text of 18 SRs was retrieved and screened against the eligibility criteria for both research questions 1 and 2.

3.2. Research Question 1: Are there systematic reviews explicitly evaluating prognostic models of non–surgical treatment outcomes for lumbar–related leg pain?

None of the SRs screened by their full text explicitly evaluated prognostic models of treatment outcomes for persons with LRLP. Table 2 presents the reasons for exclusion for each of these reviews. Further elaboration on the reasons for exclusion is presented in Appendix B. Most of the reviews included persons with LRLP but either failed to provide the percentage of this cohort within the total number of participants (Silva *et al.*, 2022; Karran *et al.*, 2017a), included less than 80% of the participants with LRLP (May & Rosedale, 2009; Kent & Keating, 2008; Haskins, Osmotherly & Rivett, 2015; Tagliaferri *et al.*, 2022; Ogbeivor & Elsabbagh, 2021), or both (Haskins, Rivett & Osmotherly, 2012; Stanton *et al.*, 2010), or did not include persons with LRLP (Fu *et al.*, 2024; Lheureux & Berquin, 2019; McIntosh *et al.*, 2018; Patel *et al.*, 2013; Almas, Parsons & Whalen, 2018). Therefore, this scoping review has determined that there

are no SRs that evaluate prognostic models of non–surgical interventions, specifically in persons suffering from LRLP.

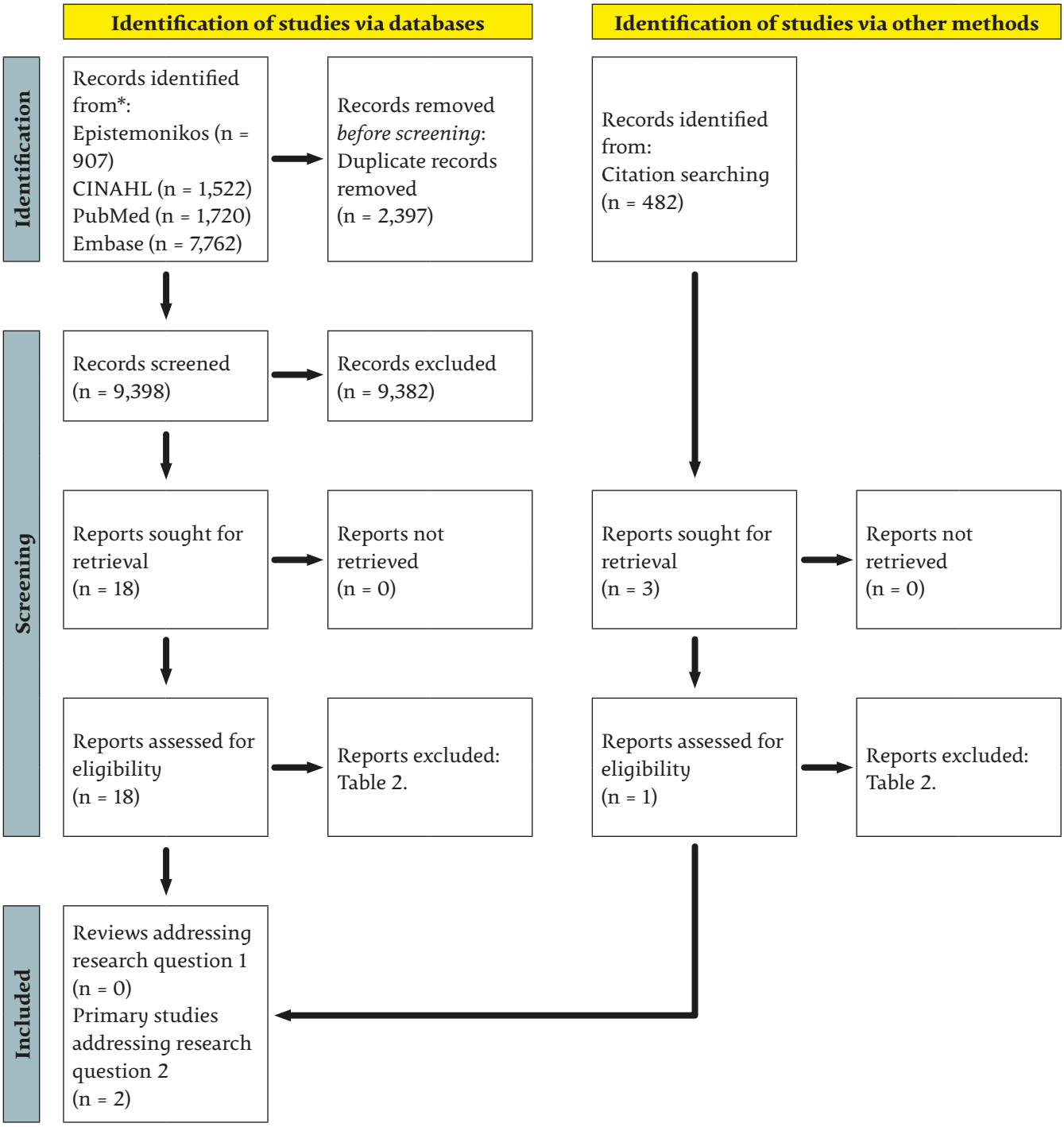

Figure 1. The PRISMA flow diagram for the scoping review

Table 2. Reasons for exclusion.

Systematic review	Reason
Fu *et al.*, (2024)	The review excluded studies evaluating persons with low back pain caused by disc herniations.
Feller *et al.*, (2024)	Evaluated prognostic models related to the surgical outcomes of lumbar–related leg pain (Fritzell, Mesterton and Hagg, 2022; Staartjes *et al.,* 2019) or diagnostic models (Stynes *et al.,* 2018).
Silva *et al.*, (2022) Karran *et al.*, (2017a)	The included primary studies either did not include persons with lumbar–related leg pain or, if they did, did not mention the proportion of such patients in respect to the total cohort; or less than 80% of the participants had lumbar–related leg pain.
Tagliaferri *et al.*, (2020)	The included studies evaluated prediction models of treatment outcomes for low back pain patients only or evaluated a model to predict who would experience recurrent lumbar disc herniations, yet this was not associated with any treatment.
Lheureux & Berquin, (2019) McIntosh *et al.*, (2018) Patel *et al.*, (2013)	None of the models included in these reviews evaluated persons with lumbar–related leg pain.
Haskins, Rivett & Osmotherly, (2012)	This review included two prognostic models; however, in one of the models the proportion of persons with lumbar–related leg pain was less than 80% of the cohort (n=28), and in the other study, the proportion of patients with lumbar–related leg pain was not provided.
Stanton *et al.*, (2010)	The included studies either excluded persons with lumbar–related leg pain or included studies with less than 80% of the participants having lumbar–related leg pain, or they did not mention the proportion of such patients in respect to the total cohort.
May & Rosedale, (2009) Kent & Keating, (2008)	The included studies had less than 80% of the participants having lumbar–related leg pain.
Haskins, Osmotherly & Rivett, (2015) Tagliaferri *et al.*, (2022) Ogbeivor & Elsabbagh, (2021)	The main aim of the reviews was to evaluate prediction models for low back pain disorders, and most of the studies included a cohort with less than 80% of the participants having lumbar–related leg pain.
King *et al.*, (2015)	No full text is available (presented only as a poster).
Chiodo & Haley, (2024)	Included only studies evaluating low back pain.
Almas, Parsons & Whalen, (2018)	The systematic review did not include participants with lumbar–related leg pain.

3.3. Research Question 2: Are there prognostic models addressing non–surgical treatment outcomes for individuals with lumbar–related leg pain?

Within the SRs which were excluded at the full–text stage for the first research question, there were three SRs (Haskins, Osmotherly & Rivett, 2015; Tagliaferri *et al.*, 2022; Ogbeivor & Elsabbagh, 2021) which included a primary study that derived or evaluated a prognostic model in a cohort composed of at least 80% of the participants having LRLP. Therefore, these primary studies (Kovacs *et al.*, 2012; Konstantinou *et al.*, 2020) were eligible for evaluation for our second research question. Since the SCOPIC trial evaluated the effectiveness of an algorithm, hence it was at the PROGRESS 4 level, we sought the paper on the development of the algorithm (Konstantinou *et al.*, 2019). Two of the SRs (Haskins, Osmotherly & Rivett, 2015; Tagliaferri *et al.*, 2022) included other prognostic models; however since these were evaluated in a cohort of less than 80% of persons with LRLP, they did not fulfil the inclusion criteria.

3.4. Characteristics of the systematic reviews

Table 3 presents an overview of the three SRs, which included the primary studies evaluating a prognostic model or CPR within a cohort of participants with LRLP. The review by Haskins, Osmotherly & Rivett, (2015) identified 30 CPRs, one of which was the multivariate predictive logistic regression model developed by Kovacs *et al.*, (2012). The review by Tagliaferri *et al.*, (2022) included 24 trials that evaluated the efficacy of classification approaches for managing LBP. Four of these trials used the SBT, and one trial (SCOPIC) included a cohort solely composed of persons with LRLP (Konstantinou *et al.*, 2020). Ogbeivor & Elsabbagh (2021) sought to determine the effectiveness of stratified care using the SBT compared to standard physiotherapy for LBP. The SR included seven trials and two health economic analyses for two of the seven included trials. One of the trials was also the SCOPIC trial (Konstantinou *et al.*, 2020).

3.5. Characteristics of the primary studies and their prognostic models

Kovacs *et al.*, (2012) aimed to develop three separate multivariate predictive logistic regression models to quantify the likelihood for a given patient to experience a clinically relevant improvement in LBP, leg pain and disability, respectively. The data was extracted from a post–marketing surveillance register for neuroreflexotherpay (NRT), which was defined as the implantation of surgical material in specific areas of the skin for up to 90 days (Urrútia *et al.*, 2004). The dependent variable for the model evaluating leg pain was the 0–10cm VAS scale. The improvement in pain was defined as any reduction in the VAS score that was higher than the minimal clinically important change (≥1.5 VAS points). Discrimination was evaluated using the areas under the receiver operating characteristic curve (AUC), while calibration was assessed using the Hosmer–Lemeshow test. Internal validation of the model was tested via 1,000 bootstrapping samples. Multiple imputation was done for missing data, resulting in five imputed data sets. Variable selection was performed for each dataset, but a variable was included in the combined model if it appeared as a predictor in at least two of the five imputed datasets. The regression coefficients were averaged using Rubin rules. A total of 4,477 participants were registered on the registry, with only 4.8% missing data for one or more variables, and only 0.07% of the patients were lost to follow–up. The data from 3,359 participants was used to develop the predictive model for LRLP. 75.2% of the participants with referred leg pain showed a clinically relevant improvement at discharge, while the rest did not. The model was composed of the following variables: had been treated with NRT (OR = 1.47; p<0.0001), previous surgery (–0.49; p<0.0001), baseline degree of disability (RMQ) (–0.05; p<0.0001), EMG (–0.52; p=0.021), baseline severity of LBP (VAS) (–0.07; p=0.013), baseline severity of referred pain (VAS) (0.21; p<0.0001). It obtained poor discrimination with an AUC of 0.655, and calibration was not reported.

The SCOPIC trial (Konstantinou *et al.*, 2020) was included in two SRs (Tagliaferri *et al.*, 2022; Ogbeivor & Elsabbagh, 2021) and it evaluated whether the SBT could provide any benefit compared to usual care in UK primary care for persons with sciatica. The SCOPIC trial was a two–parallel arm, pragmatic RCT (PROGRESS 4) within three centres in the UK, enrolling adults with a clinical diagnosis of sciatica. Patients were randomly allocated to either stratified care or usual care. Stratified care consisted of 3 risk–based groups: group 1 received brief advice and two physiotherapy sessions, group 2 received up to six physiotherapy sessions, and Group 3 was fast–tracked to MRI and spinal specialist assessment within 4 weeks of randomisation. The primary outcome

was time to first resolution of the sciatic symptoms, defined as "completely recovered" or "much better" on a 6–point Likert scale. It enrolled 238 patients in each treatment arm. There was not a statistically significant difference in the median time to symptom resolution between the two treatment arms at 4 months (p=0.66) and 12 months (p=0.22), and stratified care was deemed not cost–effective compared to usual care.

The algorithm used within the SOCPIC trial was dependent on the SBT, but included other variables being current leg pain intensity (NRS 0–10 with a cut–off >6), pain radiating below the knee (yes/no), pain interference with work or home activities (NRS 0–10 with a cut–off >6) and objective sensory loss (yes/no) subgroup (Konstantinou *et al.*, 2019). The algorithm was derived by first conducting a logistic regression analysis with the dependent variable being the patient being referred to specialist services (yes/no). The model achieved a calibration of 1.0 (95%CI 0.57 to 1.43), and an AUC of 0.695 (95%CI 0.622 to 0.768), and after conducting bootstrapping, the AUC was 0.678 (95%CI 0.674 to 0.681). At this stage, three variables obtained statistical significance within the model, being: pain interference with work or home activities (aOR 2.17; 95% CI 1.13 to 4.17), current leg pain intensity (aOR 1.17; 95%CI 1.05 to 1.31) and sensory loss (aOR 2.41; 95% CI 1.05 to 5.53). The predictive factors were discussed with various experts who added an extra variable: pain below the knee (yes/no) since this is considered the best proxy indicator of leg pain due to nerve root involvement. Furthermore, the experts added cut–off points for the impact and pain intensity scales since these thresholds for pain and functional limitations are considered reasonable and have face validity for considering early referral to specialists. Afterwards, multiple combinations, including the SBT cut–off scores for the three risk groups and the four new variables, were computed, aiming to achieve the optimal combination for onward referral to specialist services based on sensitivity, specificity, and negative and positive predictive values. The authors and experts identified the optimal criteria for prioritizing sciatica patients for spinal specialist services as either: (1) a high–risk (≥4 on the psychological subscale score) on the SBT combined with at least three of the four specified clinical characteristics, or (2) a medium–risk (3≤ on the psychological subscale score) classification on the SBT alongside the presence of all four clinical characteristics. This algorithm achieved a sensitivity of 51% (95% CI 37 to 64), a specificity of 73% (95% CI 68 to 78), a positive predictive value of 22% (95% CI 16 to 31), a negative predictive value of 91% (95% CI 87 to 94) and 30% (129 out of 429 participants) of the total sample being referred for specialist services. The positive predictive value indicated that 22% of the patients being referred to specialist services would be appropriately referred. This algorithm was based on the sensitivity value and feasibility of the spinal specialist services to handle the number of referrals.

A direct comparison between the two models is challenging due to several factors, including differences in the healthcare settings (UK vs. Spanish primary care), the dependent variables predicted by each model (dichotomized pain intensity vs. onward referral to specialist services), the PROGRESS levels used (PROGRESS 3 vs. 4), and the methodological approaches adopted (logistic regression vs. mixed methods). The model developed by Kovacs *et al.*, (2012) could be subject to bias due to potential conflicts of interest stemming from pharmaceutical company involvement, which could have influenced the reported positive effect of NRT. However, despite these findings, NRT is not recommended in any major clinical practice guidelines for LBP with or without LRLP (NICE, 2020a). Additionally, Kovacs *et al.*, (2012) dichotomized a continuous outcome (VAS 0–10), thereby reducing the statistical power of their prognostic model. In contrast, the algorithm used in the SCOPIC trial (Konstantinou *et al.*, 2020) demonstrated slightly better AUC values in comparison to the Kovacs *et al.* (2012) model (0.678 vs. 0.655) and was well calibrated, with a calibration slope of 1.0 (95% CI 0.57 to 1.43). Notably, the calibration value for the model developed by Kovacs *et al.* (2012) was not reported.

3.6. Critical appraisal

The Kovacs *et al.*, (2012) model was at high risk of bias (Table 4) since the data was obtained from a routine care registry being a marketing surveillance study on NRT, explaining why 94.8% of the participants within the study had received such therapy. The final model included the variable "treated with neuroreflexology". This variable would not be available at baseline since the patient would not have yet received treatment, making the model unusable at baseline. The bias derived from this variable in the model is reflected by the high association with the outcome (coefficient = 1.47) compared to the magnitude of the other coefficients. The information on the model's calibration was not provided. Instead, the authors provided the p–value (0.156) of the Hosmer–Lemeshow test, which does not indicate the presence nor the magnitude of any miscalibration. All

these factors lead to downgrading the risk of bias grade and applicability for this model.

The algorithm used in the SCOPIC trial is not strictly a prediction model based on statistical regression analysis, but it utilises a mixed methodology (Konstantinou *et al.*, 2019). After conducting the initial logistic regression analysis, the authors sought an iterative process involving several combinations to inflate the algorithm's performance, therefore downgrading the risk of bias grade for the outcome domain. The initial logistic regression model had an event per variable (EPV) ratio of 2.6, which would be even smaller when adding the three risk groups of the SBT. Ideally, an EPV of at least 20 is recommended in model development studies (Wolff *et al.*, 2019). The EPV ratio refers to the number of outcome events (e.g., patients experiencing an event) available per predictor variable in a prognostic model. A low EPV increases the risk of overfitting, making the model unreliable. Although a few participants had missing baseline data, the authors did not provide information on how this was tackled, for example, whether multiple imputation was used. These lead to downgrading the risk of bias grade. Figure 2 provides an overall summary of the risk of bias assessment using PROBAST and further detail is provided in Appendix C.

Table 3. Characteristics of the systematic reviews and the eligible primary studies.

Review	**Search span / Databases**	**Aims of the review**	**Primary study / Prognostic model/tool**	**Eligibility criteria pertaining to lumbar–related leg pain within the primary study**
Haskins, Osmotherly & Rivett, (2015)	From inception to July 2013. MEDLINE, EMBASE, CENTRAL, PsychINFO, CINAHL, AMED, and Index to Chiropractic Literature.	To identify prognostic forms of clinical prediction rules related to the nonsurgical management of adults with LBP and to evaluate their current stage of development.	Kovacs *et al.*, (2012) Multivariate predictive logistic regression model.	Inclusion criteria encompassed patients with low back pain, with or without leg pain, without trauma or systemic disease causes, and could read Spanish. Patients with prior unsuccessful spine surgery were eligible.

Table 3. Characteristics of the systematic reviews and the eligible primary studies.

Tagliaferri *et al.*, (2022)	From inception to June 2021. MEDLINE, EMBASE, CINAHL, the Web of Science Core Collection, CENTRAL.	To determine the efficacy of nonsurgical classification systems for treating LBP, compared to general comparators, for LBP, leg pain intensity, and disability. The secondary aim was to determine the effectiveness of treating subclasses of individuals in classification systems.	Konstantinou, *et al.*, (2020) Algorithm based on the STarT Back tool (SCOPIC trial)	Eligible patients were 18 years or older, with a clinical diagnosis of sciatica of any severity and duration, following confirmed by assessment, and had not received treatment for this condition within the last three months. Sciatica was defined by symptoms such as leg pain approximating following a dermatomal pattern, leg pain equal to or worse than back pain, leg pain aggravated by coughing or straining, dermatomal sensory changes, neurological deficits indicating nerve root compression, a positive neural tension test, or leg pain exacerbated by weight–bearing and relieved by sitting (specifically spinal stenosis).
Ogbeivor & Elsabbagh, (2021)	From 1 January 2000 to 5 July 2020. CINAHL, MEDLINE, Pedro, EMBASE, PsycINFO, CENTRAL and Web of Science.	The aim was to investigate stratified care's long–term clinical, and cost–effectiveness compared with non–stratified care to determine which approach is better for the long–term management of patients with LBP.		

CENTRAL – Cochrane Central Register of Controlled Trials

Table 4. Tabular presentation of PROBAST results.

Study	ROB				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Kovacs *et al.*, (2012)	–	–	+	–	+	–	–	–	–
Konstantinou *et al.*, (2020)	+	+	–	?	+	+	+	–	+

PROBAST = Prediction model Risk Of Bias ASsessment Tool; ROB = risk of bias.

\+ indicates low ROB/low concern regarding applicability;
– indicates high ROB/high concern regarding applicability;
? indicates unclear ROB/unclear concern regarding applicability.

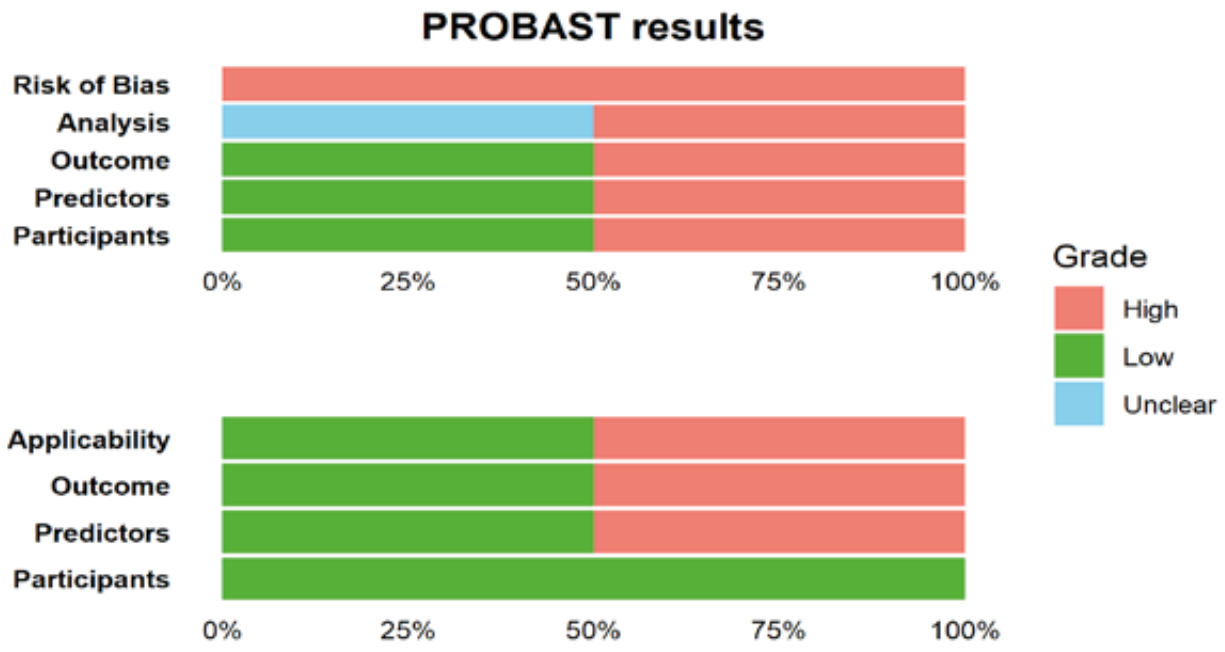

Figure 2. Summary plot of the PROBAST results.

4. Strengths and limitations of this scoping review

This scoping review demonstrates several methodological strengths. First, we conducted a comprehensive search across four databases, supplemented by forward citation searching, ensuring a broad and inclusive retrieval of relevant literature. A validated and sensitive search filter was employed for each database, except Epistemonikos, for which no validated filter currently exists. The search spanned the past 24 years, and since the review focused on SRs, their search periods potentially extended further back in time, allowing for the inclusion of a larger body of relevant literature. Moreover, the search strategy was peer–reviewed by a university librarian using the PRESS methodology, ensuring rigor and comprehensiveness. The review adhered to established reporting guidelines, with screening, data extraction, and critical appraisal of the two included prognostic models conducted independently and in duplicate to enhance reliability. Another strength lies in the investigation of the primary studies on the proportions of patients with LRLP to determine the value of the papers reviewed (as reported in Appendix B). The main limitations of this scoping review are that the search was limited to articles published in the English language, it did not search an interdisciplinary database for example, SCOPUS. Finally, this review pointed to research that needs to be conducted rather than contributing to original research.

5. Conclusion

The strengths of this scoping review provide a high level of confidence in its findings. We conclude that no SR to date has specifically evaluated prognostic models for treatment outcomes in individuals with LRLP. However, evidence indicates that apart from the two identified models (Kovacs *et al.*, 2012; Konstantinou *et al.*, 2020), various prognostic models for related leg pain have been developed (Matsudaira *et al.*, 2016; Sun *et al.*, 2023; Azharuddin *et al.*, 2022). Consequently, a SR of these models is warranted to consolidate evidence and inform clinical practice.

Funding: The research work disclosed in this publication is funded by the Tertiary Education Scholarships Scheme (Malta).

Conflict of Interest: The authors declare no conflict of interest.

Acknowledgements: The authors would like to thank Ms Agata Scicluna Derkowska, Senior Assistant Librarian at the University of Malta, for her expert assistance in developing the search strategy and reviewing the PRESS checklist.

References

Almas, M., Parsons, B. & Whalen, E. (2018). Prediction of therapeutic response to pregabalin in subjects with neuropathic pain. *Current Medical Research and Opinion*, *34*(12), 2041–2052. https://doi.org/10.1080/03007995.2018.1520694.

Azharuddin, A., Aryandono, T., Magetsari, R. & Dwiprahasto, I. (2022). Predictors of the conservative management outcomes in patients with lumbar herniated nucleus pulposus: A prospective study in Indonesia. *Asian Journal of Surgery*, *45*(1), 277–283. https://doi.org/10.1016/j.asjsur.2021.05.015.

Azimi, P., Mohammadi, H.R., Benzel, E.C., Shahzadi, S. & Azhari, S. (2015). Use of Artificial Neural Networks to Predict Recurrent Lumbar Disk Herniation. *Journal of Spinal Disorders & Techniques*, *28*(3), E161–E165. https://doi.org/10.1097/BSD.0000000000000200.

Barons, M.J., Parsons, N., Griffiths, F. & Thorogood, M. (2013). A comparison of artificial neural network, latent class analysis and logistic regression for determining which patients benefit from a cognitive behavioral approach to treatment for non–specific low back pain. *2013 IEEE Symposium on Computational Intelligence in Healthcare and E–Health (CICARE)*. https://doi.org/10.1109/cicare.2013.6583061

Beneciuk, J.M., Bishop, M.D., Fritz, J.M., Robinson, M.E., Asal, N.R., Nisenzon, A.N., *et al.* (2013). The STarT Back Screening Tool and Individual Psychological Measures: Evaluation of Prognostic Capabilities for Low Back Pain Clinical Outcomes in Outpatient Physical Therapy Settings. *Physical Therapy*, *93*(3), 321–333. https://doi.org/10.2522/ptj.20120207.

Boden, L.M., Boden, S.A., Premkumar, A., Gottschalk, M.B. & Boden, S.D. (2018). Predicting Likelihood of Surgery Before First Visit in Patients With Back and Lower Extremity Symptoms: A Simple Mathematical Model Based on More Than 8,000 Patients. *Spine*, *43*(18), 1296–1305. https://doi.org/10.1097/BRS.0000000000002603.

Chiodo, A.F. & Haley, M. (2024). Does risk stratification with a matched treatment pathway improve clinical outcomes for adults with acute back pain? A systematic review and meta–analysis. *Revista Brasileira de Fisioterapia*, *28*(5), 101116. https://doi.org/10.1016/j.bjpt.2024.101116.

Congcong, C., Yong, H.P. & Kian, C.L. (2009). A clinical prediction rule for classifying patients with low back pain who demonstrate short–term improvement with mechanical lumbar traction. *European Spine Journal*, *18*(4), 554–561. https://doi.org/10.1007/s00586-009-0909-9.

Cook, C. (2008). Potential Pitfalls of Clinical Prediction Rules. *The Journal of Manual & Manipulative Therapy*, *16*(2), 69–71. https://doi.org/10.1179/106698108790818477.

D'Agostino, R. B., Sr, Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., *et al.* (2008). General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation*, *117*(6), 743–753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579.

Feller, D., Wingbermuhle, R., Berg, B., Vigdal, ØN., Innocenti, T., Grotle, M., *et al.* (2024). Improvements Are Needed in the Adherence to the TRIPOD Statement for Clinical Prediction Models for Patients With Spinal Pain or Osteoarthritis: A Metaresearch Study. *The Journal of Pain*, *25*(11), 104624. https://doi.org/10.1016/j.jpain.2024.104624.

Field, J. & Newell, D. (2012). Relationship between STarT Back Screening Tool and prognosis for low back pain patients receiving spinal manipulative therapy. *Chiropractic & Manual Therapies*, *20*(1), 17. https://doi.org/10.1186/2045-709X-20-17.

Flynn, T., Fritz, J., Whitman, J., Wainner, R., Magel, J., Rendeiro, D., *et al.* (2002). A Clinical Prediction Rule for Classifying Patients with Low Back Pain Who Demonstrate Short–Term Improvement With Spinal Manipulation. *Spine*, *27*(24), 2835–2843. https://doi.org/10.1097/00007632-200212150-00021.

Foster, N.E., Hill, J.C., O'Sullivan, P. & Hancock, M. (2013). Stratified models of care. *Best Practice & Research Clinical Rheumatology*, *27*(5), 649–661. https://doi.org/10.1016/j.berh.2013.10.005.

Fritz, J.M., Childs, J.D. & Flynn, T.W. (2005). Pragmatic application of a clinical prediction rule in primary care to identify patients with low back pain with a good prognosis following a brief spinal manipulation intervention. *BMC Family Practice*, *6*(1), 29. https://doi.org/10.1186/1471-2296-6-29.

Fritz, J.M., Lindsay, W., Matheson, J.W., Brennan, G.P., Hunter, S.J., Moffit, S.D., *et al.* (2007). Is there a subgroup of patients with low back pain likely to benefit from

mechanical traction? Results of a randomized clinical trial and subgrouping analysis. Spine, *32*(26), E793–E800. https://doi.org/10.1097/BRS.0b013e31815d001a.

Fritzell, P., Mesterton, J. & Hagg, O. (2022). Prediction of outcome after spinal surgery—using The Dialogue Support based on the Swedish national quality register. *European Spine Journal*, *31*(4), 889–900. https://doi.org/10.1007/s00586-021-07065-y.

Fu, Y., Feller, D., Koes, B. & Chiarotto, A. (2024). Prognostic Models for Chronic Low Back Pain Outcomes in Primary Care Are at High Risk of Bias and Lack Validation–High–Quality Studies Are Needed: A Systematic Review. *The Journal of Orthopaedic and Sports Physical Therapy*, *54*(5), 1–314. https://doi.org/10.2519/jospt.2024.12081.

Gabel, C.P., Melloh, M., Yelland, M., Burkett, B. & Roiko, A. (2011). Predictive ability of a modified Örebro Musculoskeletal Pain Questionnaire in an acute/subacute low back pain working population. *European Spine Journal*, 20(3), 449–457. https://doi.org/10.1007/s00586-010-1509-4.

Gal, N., Stoicu–Tivadar, V., Andrei, D., Nemeş, D.I. & Nădăşan, E. (2014). Computer assisted treatment prediction of low back pain pathologies. *Studies in Health Technology and Informatics*, 197, 47–51. https://doi.org/10.3233/978-1-61499-389-6-47.

Geersing, G., Bouwmeester, W., Zuithoff, P., Spijker, R., Leeflang, M., Moons, K.G.M., *et al.* (2012). Search Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews. *PLoS ONE*, *7*(2), e32844. https://doi.org/10.1371/journal.pone.0032844.

George, S.Z., Bialosky, J.E. & Donald, D.A. (2005). The Centralization Phenomenon and Fear–Avoidance Beliefs as Prognostic Factors for Acute Low Back Pain: A Preliminary Investigation Involving Patients Classified for Specific Exercise. *The Journal of Orthopaedic and Sports Physical Therapy*, *35*(9), 580–588. https://doi.org/10.2519/jospt.2005.35.9.580.

Goossen, K., Hess, S., Lunny, C. & Pieper, D. (2020). Database combinations to retrieve systematic reviews in overviews of reviews: a methodological study. *BMC Medical Research Methodology*, 20(1), 1–138. https://doi.org/10.1186/s12874-020-00983-3.

Grimshaw, J. (2010). *A Guide to Knowledge Synthesis A Knowledge Synthesis Chapter.* Available at: https://cihr-irsc.gc.ca/e/41382.html

Grotle, M., Vøllestad, N.K. & Brox, J.I. (2006). Screening for Yellow Flags in First–time Acute Low Back Pain: Reliability and Validity of a Norwegian Version of the Acute Low Back Pain Screening Questionnaire. The *Clinical journal of Pain*, *22*(5), 458–467. https://doi.org/10.1097/01.ajp.0000208243.33498.cb.

Haddaway, N.R., Grainger, M.J. & Gray, C.T. (2022). Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. *Research Synthesis Methods*, 13(4), 533–545. https://doi.org/10.1002/jrsm.1563.

Hancock, M.J., Maher, C.G., Latimer, J., Herbert, R.D. & McAuley, J.H. (2009). Can rate of recovery be predicted in patients with acute low back pain? Development of a clinical prediction rule. *European Journal of Pain*, 13(1), 51–55. https://doi.org/10.1016/j.ejpain.2008.03.007.

Haskins, R., Osmotherly, P.G. & Rivett, D.A. (2015). Validation and impact analysis of prognostic clinical prediction rules for low back pain is needed: a systematic review. *Journal of Clinical Epidemiology*, *68*(7), 821–832. https://doi.org/10.1016/j.jclinepi.2015.02.003.

Haskins, R., Rivett, D.A. & Osmotherly, P.G. (2012). Clinical prediction rules in the physiotherapy management of low back pain: A systematic review. *Manual therapy*, 17(1), 9–21. https://doi.org/10.1016/j.math.2011.05.001.

Hazard, R.G., Haugh, L.D., Reid, S., McFarlane, G. & MacDonald, L. (1997). Early physician notification of patient disability risk and clinical guidelines after low back injury: A randomized, controlled trial. Spine, *22*(24), 2951–2958. https://doi.org/10.1097/00007632-199712150-00019.

Hazard, R.G., Haugh, L.D., Reid, S., Preble, J.B. & MacDonald, L. (1996). Early Prediction of Chronic Disability After Occupational Low Back Injury. *Spine*, *21*(8), 945–951. https://doi.org/10.1097/00007632-199604150-00008.

Hemingway, H., Croft, P., Perel, P., Hayden, J.A., Abrams, K., Timmis, A., *et al.* (2013). Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ*, *346*(feb05 1), 5595. https://doi.org/10.1136/bmj.e5595.

Heneweer, H., Aufdemkampe, G., Van Tulder, M.W., Kiers, H., Stappaerts, K.H. & Vanhees, L. (2007). Psychosocial variables in patients with (sub)acute low back pain: an inception cohort in primary care physical therapy in The Netherlands. *Spine*, *32*(5), 586–592. https://doi.org/10.1097/01.brs.0000256447.72623.56.

Hicks, G.E., Fritz, J.M., Delitto, A. & McGill, S.M. (2005). Preliminary Development of a Clinical Prediction Rule for Determining Which Patients With Low Back Pain Will Respond to a Stabilization Exercise Program. *Archives of Physical Medicine and Rehabilitation*, 86(9), 1753–1762. https://doi.org/10.1016/j.apmr.2005.03.033.

Hill, J.C., Whitehurst, D.G., Lewis, M., Bryan, S., Dunn, K.M., Foster, N.E., *et al.* (2011). Comparison of stratified primary care management for low back pain with current best practice (STarT Back): a randomised controlled trial. *Lancet*, 378(9802), 1560–1571. https://doi.org/10.1016/S0140-6736(11)60937-9.

Hingorani, A.D., Windt, D.A.v.d., Riley, R.D., Abrams, K., Moons, K.G.M., Steyerberg, E.W., *et al.* (2013). Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ*, 346(feb05 1), e5793. https://doi.org/10.1136/bmj.e5793.

Holland, J.L., Wilczynski, N.L. & Haynes, R.B. (2005). Optimal search strategies for identifying sound clinical prediction studies in EMBASE. *BMC Medical Informatics and Decision Making*, 5(1), 11. https://doi.org/10.1186/1472-6947-5-11.

Jellema, P., van der Windt, Daniëlle A W M, van der Horst, H.E., Stalman, W.A.B. & Bouter, L.M. (2007). Prediction of an unfavourable course of low back pain in general practice: comparison of four instruments. *British Journal of General Practice*, 57(534), 15–22.

Karran, E.L., McAuley, J.H., Traeger, A.C., Hillier, S.L., Grabherr, L., Russek, L.N., *et al.* (2017). Can screening instruments accurately determine poor outcome risk in adults with recent onset low back pain? A systematic review and meta–analysis. *BMC Medicine*, 15(1) https://doi.org/10.1186/s12916-016-0774-4.

Kavanagh, P.L., Frater, F., Navarro, T., LaVita, P., Parrish, R. & Iorio, A. (2021). Optimizing a literature surveillance strategy to retrieve sound overall prognosis and risk assessment model papers. *Journal of the American Medical Informatics Association,* 28(4), 766–771. https://doi.org/10.1093/jamia/ocaa232.

Kendell, M., Beales, D., O'Sullivan, P., Rabey, M., Hill, J. & Smith, A. (2018). The predictive ability of the STarT Back Tool was limited in people with chronic low back pain: a prospective cohort study. *Journal of Physiotherapy*, 64(2), 107–113. https://doi.org/10.1016/j.jphys.2018.02.009.

Kent, P.M. & Keating, J.L. (2008). Can we predict poor recovery from recent–onset nonspecific low back pain? A systematic review. *Manual Therapy*, 13(1), 12–28. https://doi.org/10.1016/j.math.2007.05.009.

King, P., Galvin, R., O'Sullivan & Bennett, Z. (2015). A systematic review of clinical prediction rules to aid treatment selection in musculoskeletal physiotherapy practice. *BMC Proceedings*, 9(S1), A25. https://doi.org/10.1186/1753-6561-9-S1-A25.

Kongsted, A., Andersen, C.H., Hansen, M.M. & Hestbaek, L. (2016). Prediction of outcome in patients with low back pain – A prospective cohort study comparing clinicians' predictions with those of the Start Back Tool. *Manual Therapy*, 21, 120–127. https://doi.org/10.1016/j.math.2015.06.008.

Konstantinou, K., Dunn, K.M., Van Der Windt, D., Ogollah, R., Jasani, V., Foster, N.E., Artus, M., Hay, E.M., Lewis, M., Hill, J.C. & Mallen, C. (2019). Subgrouping patients with sciatica in primary care for matched care pathways: Development of a subgrouping algorithm. *BMC Musculoskeletal Disorders*, 20(1), 313. https://doi.org/10.1186/s12891-019-2686-x

Konstantinou, K., Dunn, K.M., Ogollah, R., Lewis, M., van der Windt, D. & Hay, E.M. (2018). Prognosis of sciatica and back–related leg pain in primary care: the ATLAS cohort. *The Spine Journal*, 18(6), 1030–1040. https://doi.org/10.1016/j.spinee.2017.10.071.

Konstantinou, K., Lewis, M., Dunn, K.M., Ogollah, R., Artus, M., Hill, J.C., *et al.* (2020). Stratified care versus usual care for management of patients presenting with sciatica in primary care (SCOPiC): a randomised controlled trial. *The Lancet Rheumatology*, 2(7), e401–e411. https://doi.org/10.1016/S2665-9913(20)30099-0.

Kovacs, F.M., Seco, J., Royuela, A., Corcoll Reixach, J., & Abraira, V. (2012). Predicting the evolution of low back pain patients in routine clinical practice: results from a registry within the Spanish National Health Service. *The Spine Journal*, 12(11), 1008–1020. https://doi.org/10.1016/j.spinee.2012.10.007.

Kraemer, H.C., Frank, E. & Kupfer, D.J. (2006). Moderators of Treatment Outcomes. *JAMA*, 296(10), 1286. https://doi.org/10.1001/jama.296.10.1286.

Law, R.K.Y., Lee, E.W.C., Law, S., Chan, B.K.B., Chen, P. & Szeto, G.P.Y. (2013). The Predictive Validity of OMPQ on the Rehabilitation Outcomes for Patients with Acute and Subacute Non–Specific LBP in a Chinese Population. *Journal of Occupational Rehabilitation*, 23(3), 361–370. https://doi.org/10.1007/s10926-012-9404-y.

Lheureux, A. & Berquin, A. (2019). Comparison between the STarT Back Screening Tool and the Örebro Musculoskeletal Pain Screening Questionnaire: Which tool for what purpose? A semi–systematic review. *Annals of Physical and Rehabilitation Medicine*, 62(3), 178–188. https://doi.org/10.1016/j.rehab.2018.09.007.

Linton, S.J. & Boersma, K. (2003). Early Identification of Patients at Risk of Developing a Persistent Back Problem: The Predictive Validity of The Örebro Musculoskeletal Pain Questionnaire. The Clinical Journal of Pain, 19(2), 80–86. https://doi.org/10.1097/00002508-200303000-00002.

Matsudaira, K., Hara, N., Oka, H., Kunogi, J., Yamazaki, T., Takeshita, K., *et al.* (2016). Predictive Factors for Subjective Improvement in Lumbar Spinal Stenosis Patients with Nonsurgical Treatment: A 3–Year Prospective Cohort Study. *PLOS ONE*, 11(2), e0. https://doi.org/10.1371/journal.pone.0148584.

May, S. & Rosedale, R. (2009). Prescriptive Clinical Prediction Rules in Back Pain Research: A Systematic Review. *The Journal of Manual & Manipulative Therapy*, 17(1), 36–45. https://doi.org/10.1179/106698109790818214.

McGowan, J., Sampson, M., Salzwedel, D.M., Cogo, E., Foerster, V. & Lefebvre, C. (2016). PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of Clinical Epidemiology*, 75, 40–46. https://doi.org/10.1016/j.jclinepi.2016.01.021.

McIntosh, G., Steenstra, I., Hogg–Johnson, S., Carter, T. & Hall, H. (2018). Lack of Prognostic Model Validation in Low Back Pain Prediction Studies: A Systematic Review. *The Clinical Journal of Pain*, 34(8), 748–754. https://doi.org/10.1097/AJP.0000000000000591.

Mehling, W.E., Avins, A.L., Acree, M.C., Carey, T.S. & Hecht, F.M. (2015). Can a back pain screening tool help classify patients with acute pain into risk levels for chronic pain? *European Journal of Pain*, 19(3), 439–446. https://doi.org/10.1002/ejp.615.

Mehling, W.E., Ebell, M.H., Avins, A.L., & Hecht, F.M. (2015). Clinical decision rule for primary care patient with acute low back pain at risk of developing chronic pain. *The Spine Journal*, 15(7), 1577–1586. https://doi.org/10.1016/j.spinee.2015.03.003.

Moons, K.G.M., de Groot, J.A.H., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D.G., *et al.* (2014). Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Medicine*, 11(10), e1001744. https://doi.org/10.1371/journal.pmed.1001744 16.

Moons K.G.M., Altman, D.G., Reitsma, J.B., Ioannidis, J.P.A., Macaskill, P., Steyerberg, E.W., *et al.* (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med*, 162(1–73). https://doi.org/10.7326/M14-0698.

Morsø, L., Kent, P., Manniche, C. & Albert, H.B. (2014). The predictive ability of the STarT Back Screening Tool in a Danish secondary care setting. *European Spine* Journal, 23(1), 120–128. https://doi.org/10.1007/s00586-013-2861-y.

Newell, D., Field, J. & Pollard, D. (2015). Using the STarT Back Tool: Does timing of stratification matter? *Manual Therapy*, 20(4), 533–539. https://doi.org/10.1016/j.math.2014.08.001.

NICE (2020a). *Low back pain and sciatica in over 16s: assessment and management*. National Institute for Health and Care Excellence (NICE).

NICE (2020b). *Endorsed resource – STarT Back Screening Tool with matched treatment options.* National Institute for Health and Care Excellence (NICE). Available at: https://www.nice.org.uk/guidance/ng59/resources/endorsed-resource-start-back-screening-tool-with-matched-treatment-options-4906309933.

Nonclercq, O. & Berquin, A. (2012). Predicting chronicity in acute back pain: Validation of a French translation of the Örebro Musculoskeletal Pain Screening Questionnaire. *Annals of Physical and Rehabilitation Medicine*, 55(4), 263–278. https://doi.org/10.1016/j.rehab.2012.03.002.

Ogbeivor, C. & Elsabbagh, L. (2021). Management approach combining prognostic screening and targeted treatment for patients with low back pain compared with standard physiotherapy: A systematic review & meta–analysis. *Musculoskeletal Care*, 19(4), 436–456. https://doi.org/10.1002/msc.1541.

Parreira, P., Maher, C.G., Steffens, D., Hancock, M.J. & Ferreira, M.L. (2018). Risk factors for low back pain and sciatica: an umbrella review. *The Spine Journal*, 18(9), 1715–1721. https://doi.org/10.1016/j.spinee.2018.05.018.

Patel, S., Friede, T., Froud, R., Evans, D.W. & Underwood, M. (2013). Systematic Review of Randomized Controlled Trials of Clinical Prediction Rules for Physical Therapy in Low Back Pain. *Spine*, 38(9), 762–769. https://doi.org/10.1097/BRS.0b013e31827b158f.

Peters, M.D.J., Marnie, C., Tricco, A.C., Pollock, D., Munn, Z., Alexander, L., *et al.* (2020). Updated methodological guidance for the conduct of scoping reviews. *JBI*

Evidence Synthesis, 18(10), 2119–2126. https://doi.org/10.11124/jbies-20-00167.

Riley, R.D., Hayden, J.A., Steyerberg, E.W., Moons, K.G.M., Abrams, K., Kyzas, P.A., *et al.* (2013). Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLoS Medicine*, 10(2), e1–e1001380. https://doi.org/10.1371/journal.pmed.1001380.

Schmidt, C.O., Kohlmann, T., Pfingsten, M., Lindena, G., Marnitz, U., Pfeifer, K., *et al.* (2016). Construct and predictive validity of the German Örebro questionnaire short form for psychosocial risk factor screening of patients with low back pain. *European Spine Journal*, 25(1), 325–332. https://doi.org/10.1007/s00586-015-4196-3.

Shaw, W.S., Pransky, G. & Winters, T. (2009). The Back Disability Risk Questionnaire for Work–Related, Acute Back Pain: Prediction of Unresolved Problems at 3–Month Follow–Up. *Journal of Occupational and Environmental Medicine*, 51(2), 185–194. https://doi.org/10.1097/JOM.0b013e318192bcf8.

Shea, B.J., Reeves, B.C., Wells, G., Thuku, M., Hamel, C., Moran, J., *et al.* (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non–randomised studies of healthcare interventions, or both. *BMJ (Online)*, 358, j4008. https://doi.org/10.1136/bmj.j4008.

Silva, F.G., Costa, L.O., Hancock, M.J., Palomo, G.A., Costa, L.C. & da Silva, T. (2022). No prognostic model for people with recent–onset low back pain has yet been demonstrated to be suitable for use in clinical practice: a systematic review. *Journal of Physiotherapy*, 68(2), 99–109. https://doi.org/10.1016/j.jphys.2022.03.009.

Silva, T., Macaskill, P., Kongsted, A., Mills, K., Maher, C.G. & Hancock, M.J. (2019). Predicting pain recovery in patients with acute low back pain: Updating and validation of a clinical prediction model. *European Journal of Pain*, 23(2), 341–353. https://doi.org/10.1002/ejp.1308.

Staartjes, V.E., de Wispelaere, M.P., Vandertop, W.P. & Schröder, M.L. (2019). Deep learning–based preoperative predictive analytics for patient–reported outcomes following lumbar discectomy: feasibility of center–specific modeling. *The Spine Journal*, 19(5), 853–861. https://doi.org/10.1016/j.spinee.2018.11.009.

Stanton, T.R., Hancock, M.J., Maher, C.G. & Koes, B.W. (2010). Critical Appraisal of Clinical Prediction Rules That Aim to Optimize Treatment Selection for Musculoskeletal Conditions. *Physical Therapy*, 90(6), 843–854. https://doi.org/10.2522/ptj.20090233.

Steyerberg, E.W., Moons, K.G.M., van der Windt, D.A., Hayden, J.A., Perel, P., Schroter, S., *et al.* (2013). *Prognosis Research Strategy (PROGRESS) 3*: Prognostic Model Research. *PLoS Medicine*, 10(2), e1–e1001381. https://doi.org/10.1371/journal.pmed.1001381.

Stynes, S., Konstantinou, K., Ogollah, R., Hay, E.M. & Dunn, K.M. (2018). Clinical diagnostic model for sciatica developed in primary care patients with low back–related leg pain. *PloS One*, 13(4), e0191852. https://doi.org/10.1371/journal.pone.0191852.

Sun, D., Liu, Y., Luo, D., Wu, Y., Yan, Z., Liang, Y., *et al.* (2023). A multidimensional nomogram combining clinical factors and imaging features to predict 1–year recurrence of low back pain with or without radicular pain after spinal manipulation/mobilization. *Chiropractic & Manual Therapies*, 31(1), 27. https://doi.org/10.1186/s12998-023-00500-5.

Tagliaferri, S.D., Angelova, M., Zhao, X., Owen, P.J., Miller, C.T., Wilkin, T., *et al.* (2020). Artificial intelligence to improve back pain outcomes and lessons learned from clinical classification approaches: three systematic reviews. *NPJ Digital Medicine*, 3(1), 93. https://doi.org/10.1038/s41746-020-0303-x.

Tagliaferri, S.D., Mitchell, U.H., Saueressig, T., Owen, P.J., Miller, C.T. & Belavy, D.L. (2022). Classification Approaches for Treating Low Back Pain Have Small Effects That Are Not Clinically Meaningful: A Systematic Review With Meta–analysis. *The Journal of Orthopaedic and Sports Physical Therapy*, 52(2), 67–84. https://doi.org/10.2519/jospt.2022.10761.

Traeger, A.C., Henschke, N., Hübscher, M., Williams, C.M., Kamper, S.J., Maher, C.G., *et al.* (2016). Estimating the Risk of Chronic Pain: Development and Validation of a Prognostic Model (PICKUP) for Patients with Acute Low Back Pain. *PLoS Medicine*, 13(5), e1002019. https://doi.org/10.1371/journal.pmed.1002019.

Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., *et al.* (2018). PRISMA Extension for Scoping Reviews (PRISMA–ScR): Checklist and Explanation. *Annals of Internal Medicine*, 169(7), 467–473. https://doi.org/10.7326/M18-0850.

Truchon, M., Schmouth, M., Côté, D., Fillion, L., Rossignol, M. & Durand, M. (2012). Absenteeism Screening Questionnaire (ASQ): A New Tool for Predicting Long–term Absenteeism Among Workers with Low Back

Pain. *Journal of Occupational Rehabilitation*, *22*(1), *27–50*. https://doi.org/10.1007/s10926-011-9318-0.

Tsang Chi Chung, R., Lau Sau Ying, J., Kwong So Fong, S., So Ming Loi, E., Law Ka Yee, R., Wong Fu Yan, T., *et al.* (2017). Reliability, Construct and Predictive Validity of the Hong Kong Chinese Orebro Musculoskeletal Pain Screening Questionnaire. *Journal of Occupational Rehabilitation*, *27*(4), *584–592*. https://doi.org/10.1007/s10926-016-9689-3.

Tsang, R.C.C., Lee, E.W.C., Lau, J.S.Y., Kwong, S.S.F., So, E.M.L., Wong, T.F.Y., *et al.* (2019). Development of a short form of the Hong Kong Chinese Orebro musculoskeletal pain screening questionnaire. *Hong Kong Physiotherapy Journal*, *39*(1), *57–66*. https://doi.org/10.1142/S1013702519500057.

Turner, J.A., Shortreed, S.M., Saunders, K.W., LeResche, L., Berlin, J.A. & Korff, M.V. (2013). Optimizing prediction of back pain outcomes. *Pain*, *154*(8), 1391–1401. https://doi.org/10.1016/j.pain.2013.04.029.

Urrútia, G., Burton, A.K., Morral Fernández, A., Bonfill Cosp, X. & Zanoli, G. (2004). Neuroreflexotherapy for non–specific low–back pain. *Cochrane Database of Systematic Reviews*, 2011(2), CD003009. https://doi.org/10.1002/14651858.CD003009.pub2.

Walker–Dilks, C., Wilczynski, N.L. & Haynes, R.B. (2008). Cumulative Index to Nursing and Allied Health Literature search strategies for identifying methodologically sound causation and prognosis studies. *Applied Nursing Research*, *21*(2), *98–103*. https://doi.org/10.1016/j.apnr.2006.04.004.

Williams, C.M., Hancock, M.J., Maher, C.G., McAuley, J.H., Lin, C.W.C. & Latimer, J. (2014). Predicting rapid recovery from acute low back pain based on the intensity, duration and history of pain: A validation study. *European Journal of Pain*, *18*(8), 1182–1189. https://doi.org/10.1002/j.1532-2149.2014.00467.x.

Wolff, R.F., Moons, K.G.M., Riley, R.D., Whiting, P.F., Westwood, M., Collins, G.S., *et al.* (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, *170*(1), *51–58*. https://doi.org/10.7326/M18-1376.

Wolff, S.V., Willburger, R., Hallner, D., Rusu, A.C., Rusche, H., Schulte, T., *et al.* (2018). Avoidance–Endurance Fast–Screen (AE–FS). Schmerz, *32*(4), *283–292*. https://doi.org/10.1007/s00482-018-0310-x.

Appendix A: Search strings

5.1. PubMed

Search	Query	Results
#12	Search: (#6 AND #5) Filters: Meta-Analysis, Systematic Review, Humans, English, from 2000–2024 Sort by: Most Recent	1,720
#11	Search: (#6 AND #5) Filters: Meta-Analysis, Systematic Review, Humans, from 2000–2024 Sort by: Most Recent	1,748
#10	Search: (#6 AND #5) Filters: Meta-Analysis, Systematic Review, from 2000–2024 Sort by: Most Recent	2,016
#9	Search: (#6 AND #5) Filters: Meta-Analysis, from 2000–2024 Sort by: Most Recent	827
#8	Search: (#6 AND #5) Filters: from 2000–2024	33,531
#7	Search: #6 AND #5	37,015
#6	Search: #1 OR #2 OR #3 OR #4	154,049
#5	Search: (((((((Validat* OR Predict*[Title] OR Rule*) OR (Predict* AND (Outcome* OR Risk* OR Model*)) OR ((History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor*) AND (Predict* OR Model* OR Decision* OR Identif* OR Prognos*)) OR (Decision* AND (Model* OR Clinical* OR Logistic Models/)) OR (Prognostic AND (History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor* OR Model*))) OR ("Stratification" OR "ROC Curve" [Mesh] OR "Discrimination" OR "Discriminate" OR "c-statistic" OR "c statistic" OR "Area under the curve" OR "AUC" OR "Calibration" OR "Indices" OR "Algorithm" OR "Multivariable"))))))	7,578,220
#4	Search: (spin* pain[Title/Abstract]) OR (Low Back Pain[MeSH Terms]) OR (low back pain[Title/Abstract])	47,463
#3	Search: (neuropathic pain [Title/Abstract]) OR (Neuralgia[MeSH Terms])	42,265
#2	Search: (Recess stenosis[Title/Abstract]) OR (Recess stenoses[Title/Abstract]) OR ("spinal stenoses"[Title/Abstract:~2]) OR ("spinal stenosis"[Title/Abstract:~2]) OR ("neurogenic claudication"[Title/Abstract:~2]) OR ("lumbar stenoses"[Title/Abstract:~2]) OR ("lumbar stenosis"[Title/Abstract:~2]) OR (Spinal Stenosis[MeSH Terms])	12,816
#1	Search: (Intervertebral Disc Displacement[MeSH Terms]) OR (Sciatic Neuropathy[MeSH Terms]) OR (Radicul*[Title/Abstract]) OR (Radiculopathy[MeSH Terms]) OR (Sciatic*[Title/Abstract]) OR (Sciatica[MeSH Terms])	72,609

5.2. Embase

No.	Query	Results
#31	#30 AND 'human'/de	7,762
#30	#29 AND (2000:py OR 2001:py OR 2002:py OR 2003:py OR 2004:py OR 2005:py OR 2006:py OR 2007:py OR 2008:py OR 2009:py OR 2010:py OR 2011:py OR 2012:py OR 2013:py OR 2014:py OR 2015:py OR 2016:py OR 2017:py OR 2018:py OR 2019:py OR 2020:py OR 2021:py OR 2022:py OR 2023:py OR 2024:py OR 2025:py) AND ('meta analysis'/de OR 'systematic review'/de)	7,959
#29	#24 AND (#25 OR #26 OR #27)	73,553
#28	predict*:ti,ab OR 'methodology'/exp OR validat*:ti,ab	10,833,771
#27	validat*:ti,ab	1,182,099
#26	'methodology'/exp	8,259,141
#25	predict*:ti,ab	2,985,802
#24	#7 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23	251,383
#23	'low back pain'/mj OR 'spinal pain'/exp OR 'spin* pain':ti,ab OR 'low* back pain':ti,ab	65,531
#22	'low* back pain':ti,ab	52,918
#21	'spin* pain':ti,ab	3,802
#20	'spinal pain'/exp	4,508
#19	'low back pain'/mj	33,127
#18	'neuropathic pain'/exp OR 'neuralgia'/de OR 'neuropathic pain':ti,ab	65,100
#17	'neuropathic pain':ti,ab	39,787
#16	'neuralgia'/de	11,282
#15	'neuropathic pain'/exp	45,553
#14	#8 OR #9 OR #10 OR #11 OR #12 OR #13	20,980
#13	(neurogenic NEAR/2 claudication):ti,ab	1,337
#12	(lumbar NEAR/2 stenos?s):ti,ab	6,724
#11	(spinal NEAR/2 stenos?s):ti,ab	11,091
#10	'recess stenos?s':ti,ab	298
#9	'neurogenic claudication'/exp	248
#8	'vertebral canal stenosis'/exp	18,034
#7	#1 OR #2 OR #3 OR #4 OR #5 OR #6	127,556
#6	'intervertebral disk hernia'/exp	32,258
#5	'sciatica'/exp	3,463

No.	Query	Results
#4	radicul*:ti,ab	25,328
#3	'radiculopathy'/exp	48,479
#2	sciatic*:ti,ab	41,316
#1	'sciatic neuropathy'/exp	9,135

5.3. Epistemonikos

#	Query	Date
#4	The below search, dates from 2000 to 2025, filters: systematic review	907
#3	#1 AND #2	07-10-2024 03:32:01 +02:00
#2	(title:("prediction model") OR abstract:("prediction model")) OR (title:(predict*) OR abstract:(predict*)) OR (title:("prognostic model") OR abstract:("prognostic model")) OR (title:(prognos*) OR abstract:(prognos*)) OR (title:(model*) OR abstract:(model*)) OR (title:("ROC curve") OR abstract:("ROC curve")) OR (title:(discriminat*) OR abstract:(discriminat*)) OR (title:("c-statistic") OR abstract:("c-statistic")) OR (title:("c statistic") OR abstract:("c statistic")) OR (title:("area under the curve") OR abstract:("area under the curve")) OR (title:(AUC) OR abstract:(AUC)) OR (title:(calibration) OR abstract:(calibration)) OR (title:(algorithm) OR abstract:(algorithm)) OR (title:(multivariable) OR abstract:(multivariable))	07-10-2024 03:30:15 +02:00
#1	(title:(sciatic*) OR abstract:(sciatic*)) OR (title:(radicul*) OR abstract:(radicul*)) OR (title:("disc herniation") OR abstract:("disc herniation")) OR (title:("spinal stenosis") OR abstract:("spinal stenosis")) OR (title:("neurogenic claudication") OR abstract:("neurogenic claudication")) OR (title:("recess stenosis") OR abstract:("recess stenosis")) OR (title:("lumbar stenosis") OR abstract:("lumbar stenosis")) OR (title:("spinal stenosis") OR abstract:("spinal stenosis")) OR (title:("spinal pain") OR abstract:("spinal pain")) OR (title:("low back pain") OR abstract:("low back pain")) OR (title:("neuropathic pain") OR abstract:("neuropathic pain")) OR (title:("spine pain") OR abstract:("spine pain"))	07-10-2024 03:23:41 +02:00

5.4. CINAHL

#	Query	Limiters/Expanders	Last Run Via	Results
S30	S28 OR S29	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,522

#	Query	Limiters/Expanders	Last Run Via	Results
S29	S24 AND S25	Limiters – Publication Date: 20000101–20241231; Human; Publication Type: Meta Analysis Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	504
S28	S24 AND S25	Limiters – Publication Date: 20000101–20241231; Human; Publication Type: Systematic Review Expanders – Apply equivalent subjects Narrow by Language0: – english Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,448
S27	S24 AND S25	Limiters – Publication Date: 20000101–20241231 Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	32,964
S26	S24 AND S25	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	35,482
S25	S21 OR S22 OR S23	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	3,029,634
S24	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S19 OR S20	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	64,537
S23	TI outcome OR AB outcome OR MW outcome	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	907,663

#	Query	Limiters/Expanders	Last Run Via	Results
S22	TI diagnos* OR AB diagnos* OR MW diagnos*	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,264,140
S21	(MH "Study Design+")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,712,679
S20	TI "spin* pain" OR AB "spin* pain"	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,215
S19	TI "lumbar pain" OR AB "lumbar pain"	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	471
S17	TI lumbar pain OR AB "low* backpain"	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,135
S16	TI "low* backpain" OR AB "low* backpain"	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	6
S15	TI "low* back pain" OR AB "low* back pain"	Expanders – Apply equivalent subjects Search modes – Find all my search terms	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	21,745
S14	(MH "Back Pain")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	12,238

#	Query	Limiters/Expanders	Last Run Via	Results
S13	(MH "Low Back Pain")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	23,254
S12	(MH "Neuralgia")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	5,223
S11	TI "neuropathic pain" OR AB "neuropathic pain"	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	7,822
S10	TI "neurogenic claudication" OR AB "neurogenic claudication"	Expanders – Apply equivalent subjects Search modes – Find all my search terms	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	290
S9	TI lumbar Stenos?s OR AB Lumbar Stenos?s	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	2,430
S8	TI Recess Stenos?s OR AB Recess Stenos?s	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	78
S7	TI Spinal Stenos?s OR AB Spinal Stenos?s	Expanders – Apply equivalent subjects Search modes – Find all my search terms	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	3,700
S6	(MH "Spinal Stenosis")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	3,211

#	Query	Limiters/Expanders	Last Run Via	Results
S5	(MH "Intervertebral Disk Displacement")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	5,054
S4	TI Radicul* OR AB Radicul*	Expanders – Apply equivalent subjects Search modes – Find all my search terms	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	4,920
S3	(MH "Radiculopathy")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	2,625
S2	(MH "Sciatica")	Expanders – Apply equivalent subjects Search modes – Proximity	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	1,825
S1	TI sciatic* OR AB sciatic*	Expanders – Apply equivalent subjects Search modes – Find all my search terms	Interface – EBSCOhost Research Databases Search Screen – Advanced Search Database – CINAHL Complete	4,937

Appendix B:
Research Question 1 – Further elaboration on the reasons for exclusion

Author	Reason
Fu *et al.*, (2024)	Studies had to include a minimum of 75% of their participants with chronic low back pain. The review excluded studies evaluating persons with LBP caused by disc herniations.
Feller *et al.*, (2024)	The review included prognostic models related to leg pain evaluated outcomes following surgery (Fritzell, Mesterton & Hagg, 2022; Staartjes *et al.*, 2019) or diagnostic models (Stynes *et al.*, 2018).

Author	Reason
Silva *et al.*, (2022)	Gabel *et al.*, (2011) evaluated the Original Orebo Musculoskeletal Pain Questionnaire, but it did not include radicular or sciatica or leg pain in the validation study.
	Jellema *et al.*, (2007) evaluated the Orebo, Low Back Pain Perception Scale (LBPPS) and a prediction rule developed by the authors themselves, but it did not mention radicular or sciatica or leg pain.
	Context: The Da Silva model was developed from the Hancock CPR. Williams *et al.*, (2014) externally validated the Hancock CPR (Hancock *et al.*, 2009). Persons with sciatica were included in the validation but not in the development sample. The authors acknowledged that the prediction rule may have limited applicability to persons with sciatica or radicular pain. The authors did not mention the proportion of persons with sciatica or leg pain. In Silva *et al.*, (2019) the authors did not mention the proportion of participants with radicular or sciatica or leg pain.
	Hancock *et al.*, (2009) looked at predicting recovery in persons with acute low back pain with or without leg pain using the Hancock CPR; however, the proportion of patients with leg pain was not mentioned.
	Both Hazard *et al.*, (1996) and Hazard *et al.*, (1997) evaluated the Vermont Disability Prediction Questionnaire, but the authors did not mention the proportion of persons with radicular or sciatica or leg pain.
	Heneweer *et al.*, (2007) evaluated the Acute Low Back Pain Screening Questionnaire (ALBPSQ), but the authors did not mention the proportion of participants with radicular or sciatica or leg pain.
	Kongsted *et al.*, (2016) compared the SBT with clinicians' expectations. They included participants with non-specific LBP or lumbar nerve root involvement (based on usual clinical practice for diagnostic triage). However, the authors did not mention the proportion of participants with radicular, sciatic or leg pain.
	Law *et al.*, (2013) evaluated the Orebo Musculoskeletal questionnaire, but only 39.0% of the participants had back pain with leg pain.
	Mehling *et al.*, (2015) evaluated the SBT, but the authors did not mention the proportion of participants with radicular, sciatica or leg pain.
	In Mehling *et al.*, (2015), only 27% of the participants had sciatic pain below the knee.
	Traeger *et al.*, (2016) evaluated the PICKUP model, but only 24% of the participants had leg pain in the development sample, and 19% had leg pain in the external validation sample.
	Truchon *et al.*, (2012) evaluated the Absenteeism Screening Questionnaire (ASQ), but the authors did not mention radicular or sciatica or leg pain.
	Both Tsang *et al.*, (2019) and Tsang Chi Chung *et al.*, (2017) evaluated the Hong Kong Chinese version of the Orebo Musculoskeletal pain screening questionnaire.
	Wolff *et al.*, (2018) evaluated the Avoidance-Endurance Fast-Screen (AE-FS) but the manuscript did not mention radicular, sciatic or leg pain.

Author	Reason
Tagliaferri *et al.*, (2020)	Azimi *et al.*, (2015) evaluated a model to predict who will experience recurrent lumbar disc herniations, yet this was not associated with any treatment.
	Barons *et al.*, (2013) evaluated a prognostic model of treatment outcomes (cognitive behavioural therapy) but only in persons with non-specific low back pain.
	Gal *et al.*, (2014) developed a prediction model of treatments for low back pain patients only.
Lheureux & Berquin, (2019) McIntosh *et al.*, (2018) Patel *et al.*, (2013)	None of the models evaluated persons with lumbar-related leg pain.

Author	Reason
Karran *et al.*, (2017a)	Kongsted *et al.* (2016) compared the SBT with clinicians' expectations. They included participants with non-specific LBP or lumbar nerve root involvement (based on usual clinical practice for diagnostic triage). However, the authors did not mention the proportion of participants with radicular, sciatic, or leg pain.
	In Beneciuk *et al.*, (2013), 66.4% of the participants had LBP with leg pain radiation.
	In Field & Newell, (2012), 36.6% (n = 404) of the participants had radiating leg pain with LBP.
	In the SBT derivation study (n=851), only 62.3% of the participants had radiation pain in the leg, and only 32.1% had radiating pain below the knee (Hill *et al.*, 2011).
	Newell, Field & Pollard, (2015) evaluated the SBT but only 45.4% of the participants (n=749) had radiation pain in the leg.
	Gabel *et al.*, (2011) evaluated the Original Orebro Musculoskeletal pain questionnaire but did not mention radicular, sciatica, or leg pain in the validation study.
	Law *et al.*, (2013) evaluated the Orebo Musculoskeletal pain questionnaire, but only 39.0% of the participants had back pain with leg pain.
	Nonclercq & Berquin, (2012) evaluated the French version of the Orebo musculoskeletal pain questionnaire, but there is no mention of radicular or sciatica or leg pain.
	Schmidt *et al.*, (2016) evaluated the German Orebo Musculoskeletal pain questionnaire, but there was no mention of radicular, sciatic or leg pain.
	Heneweer *et al.*, (2007) evaluated the Acute Low Back Pain Screening Questionnaire (ALBPSQ) but did not mention radicular, sciatic or leg pain.
	Within Grotle, Vøllestad and Brox, (2006), in the acute LBP group, only 20% of the participants had LBP with leg pain, and 56% in the chronic LBP group had leg pain.
	Williams *et al.*, (2014) sought to externally validate the Hancock CPR (Hancock *et al.*, 2009). Persons with sciatica were included in the validation but not in the development sample. The authors acknowledge that the prediction rule may have limited applicability to persons with sciatica or radicular pain. The proportion of persons with sciatica or leg pain was not provided.
	Truchon *et al.*, (2012) evaluated the Absenteeism Screening Questionnaire (ASQ), but the manuscript did not mention radicular pain, sciatic or leg pain.
	Jellema *et al.*, (2007) evaluated the Orebo, Low Back Pain Perception Scale (LBPPS) and a prediction rule developed by the authors themselves. The manuscript did not mention radicular pain, sciatic or leg pain.
	Both Hazard *et al.*, (1996) and Hazard *et al.*, (1997) evaluated the Vermont Disability Prediction Questionnaire, but the manuscript did not mention radicular pain or sciatica or leg pain.
	Turner *et al.*, (2013) evaluated the Chronic pain risk score. Only 21.4% of the participants had pain below the knee.
	Shaw, Pransky & Winters, (2009) evaluated the Back Disability Risk Questionnaire. Only 7.0% of the participants had leg pain radiating below the knee.

Author	Reason
Haskins, Rivett & Osmotherly, (2012)	This review included 25 CPRs on the physiotherapy management of LBP, two of which were prognostic models. George, Bialosky & Donald, (2005) evaluated the data from 28 participants after screening 202 participants. 21 out of 28 participants had both leg and back pain. The study evaluated whether the centralisation phenomenon and fear avoidance belief predict pain and disability at 6 months in persons with acute LBP/leg pain after attending 4 weeks of physiotherapy sessions
	Hancock *et al.*, (2009) looked at predicting recovery in persons with acute LBP with or without leg pain; however, the proportion of patients with leg pain was not provided.
Stanton *et al.*, (2010)	Congcong, Yong & Kian, (2009) included 129 patients with LBP who were referred for physiotherapy. All participants had a diagnosis related to the lumbosacral spine and had a chief complaint of pain and/or numbness in the lumbar spine, buttock, and/or lower extremity. However, only 24.8% of the participants had neurological involvement and 35.7% had pain radiation below the knee.
	Flynn *et al.*, (2002) excluded persons with signs consistent with nerve root compression (positive straight leg raise at 45°, or diminished lower extremity strength, sensation, or reflexes). The authors developed the original 5-item rule.
	Fritz, Childs & Flynn, (2005) evaluated the 2-item version of the Flynn rule (mentioned above). Participants had a primary complaint of LBP with or without referral into the lower extremity, but the proportion of patients with leg pain was not reported.
	Hicks *et al.*, (2005) included fifty-four patients with complaints of LBP with or without leg pain. In the abstract, it mentioned that persons with non-radicular LBP were studied. They excluded persons with 2 or more signs of nerve root compression: diminished lower-extremity strength, sensation, or reflexes.
May & Rosedale, (2009)	Fritz *et al.*, (2007) included participants aged between 18 and 60 years, with pain and/or numbness extending distal to the buttock in the past 24 hours, Oswestry score 30%, signs of nerve root compression (positive straight leg raise (reproduction of symptoms at <45°), or reflex, sensory, or muscle strength deficit). However, only 76.5% (49/64) of the patients had symptoms distal to the knee.
Kent & Keating, (2008)	It included studies with less than 15% of participants presenting with neuro-compressive symptoms.
Chiodo & Haley, (2024)	The included primary studies either evaluated only cases of LBP or excluded cases with radiculopathy.

Appendix C
PROBAST Assessment

Kovacs *et al.*, (2012) model		
Step 1 – All prediction models of treatment outcomes for lumbar-related leg pain		
Step 2 – Development only		
Step 3		
Domain 1. Participants.		
1.1	PN	Routine care registries
1.2	PN	Pg. 1011, second paragraph. This registry was a neuroreflexology therapy (NRT) marketing surveillance study. 94.8% of the participants had NRT (table 1), while only 15% had physiotherapy. This does not reflect usual clinical care for such conditions.
ROB – high risk of bias		
Applicability – Low (they match the review question)		
Domain 2. Predictors.		
2.1	PN	Since the data came from routine care data registry.
2.2	NI	However, predictors were measured a long time (3 months) before the outcome occurred, so they were blinded to the outcome. Therefore, the domain can still be rated as low risk of bias.
2.3	N	It included "have been treated with neu-reflexology." The treatment is not available at the time the model would be applied, making the model unusable. Also, this variable (treated with NRT) has a very high association with the outcome (coefficient = 1.47) compared to the strength of the other coefficients.
ROB – high risk of bias		
Applicability – low concern. However, the MRI scanner might not be available in the 7 primary care centres.		
Domain 3. Outcome.		
3.1	Y	They used VAS for LBP and leg pain and RMQ
3.2	Y	The standard outcome definition was VAS leg pain.
3.3	Y	They included baseline leg pain, which is partially independent of the outcomes.
3.4	PY	Outcome definition and determination were the same for all participants.
3.5	Y	The predictors in the final model were determined by independent statisticians who had no contact with the clinicians.
3.6	Y	Justified reasons for the 3-month follow-up period.
ROB – low risk of bias.		
Applicability – low concern.		

Domain 4. Analysis		
4.1	PY	Candidate predictors included in the model: Sex Age Baseline severity LBP Baseline severity leg pain Roland Morris Duration of current episodes – acute, subacute, chronic. Employment status – passive, working, receiving financial aid Recruitment setting – primary, specialised History of lumbar surgery Diagnosis of FBSS Diagnostic tests undertaken at the moment during the study period – x-ray, CT, MRI, emg, other Findings in imaging procedures – disc degeneration, facet jt, scoliosis, spondylolisthesis, spinal stenosis, annular tears, disc protrusion, disc herniation, lumbarisation of S1, sacralization of L5, other findings, no findings Treatments used – drugs, other, PT, NRT, surgery. 37 variables for 833 events. Therefore 22.5 events per variable (EPV).
4.2	Y	All continuous predictors were kept as continuous numerical variables.
4.3	PY	Despite being a routine care registry, the model included data from all the participants with leg pain.
4.4	PY	The authors used multiple imputations and included variables that were selected as predictors in at least 2 of the 5 imputed datasets. Rubin rules were used to average regression coefficients.
4.5	Y	Predictors were not selected on the basis of univariable analysis prior to multivariable modelling.
4.6	PY	For patients with subsequent episodes of LBP, it was decided that only the data from the first episode would be analysed. Patients with a baseline score below the MCIC for a given variable, except those who worsened at discharge, were excluded from the analysis.
4.7	N	Discrimination was provided for the model on leg pain as the AUC = 0.655. However, the measure of calibration was not provided. Instead, they provided the p-value (0.156) of the Hosmer-Lemeshow test as the only measure to assess the calibration of the model. However, this p-value indicates neither the presence nor the magnitude of nay miscalibration.
4.8	PY	The sample size was large (n=3,359), the model had an EPV of 22.5, and the authors conducted bootstrapping.
4.9	Y	Both predictors and coefficients correspond to the reported results of the multivariable regression model.
ROB – High. No actual value of calibration is provided.		

Methodological summary: The authors first conducted univariate analysis followed by multivariate analyses (providing discrimination and calibration and performing bootstrapping). The predictive factors in the multivariate analysis were discussed with experts. The algorithm included the StarT back and the variables included in the model (providing sensitivity, specificity, and likelihood ratios). Hence, they did a binary logistic regression model (whether or not people were referred to specialist services) and then integrated these variables into an algorithm incorporating the STarT back tool. Therefore, instead of individualised predictions, the algorithm provides group-based predictions, which are less precise for the individual and assume that the entire group has similar prognosis.

Algorithm used within SCOPIC trial (Konstantinou *et al.*, 2020; Konstantinou *et al.*, 2019)		
Step 1 – All prediction models of treatment outcomes for lumbar-related leg pain		
Step 2 – Development only		
Step 3		
Domain 1. Participants.		
1.1	PY	The data was taken from the ATLAS, which was a prospective cohort study and retrospectively analysed.
1.2	Y	It included patients with back and leg pain who were consulting by their GP.
ROB – low		
Applicability – Low (they match the review question)		
Domain 2. Predictors.		
2.1	Y	It is mentioned in the Additional file 1.
2.2	NI	Since this is a retrospective analysis, the domain can still be rated as low risk of bias.
2.3	Y	All variables were available at the time the model is intended to be used.
ROB – low. Although 2.2 is rated as NI, the algorithm developed used retrospective cohort data, and hence, it can still be rated as low risk of bias.		
Applicability – low concern.		
Domain 3. Outcome.		
3.1	Y	The outcome in the model was binary, i.e. referral or no referral to specialist services.
3.2	N	Step 2c: Identifying patients for fast-track referral to spinal specialist services: algorithm design. In this step of the iterative process, we investigated a number of possibilities in terms of combinations of factors from the clinical assessment and information on risk of poor prognosis, using the STarT Back Tool score, for identifying which patients with sciatica to refer or 'fast-track' to spinal specialist services.
3.3	Y	Predictors and outcomes were unrelated.
3.4	Y	
3.5	Y	
3.6	Y	The time interval was 1 year.
ROB – high.		
Applicability – low concern. The outcome definition is still within the review's eligibility criteria since epidural steroid injections and multidisciplinary rehabilitation can be provided within specialist services.		

Domain 4. Analysis		
4.1	PN	57 out of 429 were referred to the specialist services. They included a total of 22 candidate predictors. EPV = 2.6 + StarT back variables.
4.2	PY	Continuous variables were handled appropriately. The risk categories of the StarT back tool were originally validated in primary care LBP cases, but 32% of patients had pain radiating below the knee.
4.3	PY	Patients at low risk were excluded from the analysis, including the STarT back and for onward referral. This is justified since only 1 out of 57 patients was referred to specialist services.
4.4	NI	Table 1 of the article mentions that a small proportion has missing data for some baseline characteristics, but we are not given information about how this is tackled in the model.
4.5	PY	Univariate analysis was done, but the choice of predictors was not based on it.
4.6	NI	There is no mention of complexities, apart from Table 2, item b (in the footer).
4.7	PY	Discrimination (AUC=0.695), which after bootstrapping became 0.678. The calibration slope was 1.0 (0.57 to 1.43). The authors then provide information on the overall algorithm's sensitivity, specificity, and positive and negative likelihood ratio.
4.8	Y	Bootstrapping was done for the model.
4.9	NI	The coefficients and variables in the final model do not account for the entire list of variables included in the final algorithm.
ROB – Unclear.		